

## Review

# Artificial Vision Systems for Mobility Impairment Detection: Integrating Synthetic Data, Ethical Considerations, and Real-World Applications

Santiago Felipe Luna-Romero <sup>1,\*</sup> , Mauren Abreu de Souza <sup>1</sup>  and Luis Serpa Andrade <sup>2</sup> 

<sup>1</sup> Graduate Program on Health Technology, Pontifícia Universidade Católica do Paraná, Curitiba 80215-901, Brazil; mauren.souza@pucpr.br

<sup>2</sup> Grupo de Investigación en Hardware Embebido Aplicado (GIHEA), Universidad Politécnica Salesiana, Cuenca 010105, Ecuador; lserpa@ups.edu.ec

\* Correspondence: santiago.romero@pucpr.edu.br

**Abstract:** Global estimates suggest that over a billion people worldwide—more than 15% of the global population—live with some form of mobility disability, underscoring the pressing need for innovative technological solutions. Recent advancements in artificial vision systems, driven by deep learning and image processing techniques, offer promising avenues for detecting mobility aids and monitoring gait or posture anomalies. This paper presents a systematic review conducted in accordance with ProKnow-C guidelines, examining key methodologies, datasets, and ethical considerations in mobility impairment detection from 2015 to 2025. Our analysis reveals that convolutional neural network (CNN) approaches, such as YOLO and Faster R-CNN, frequently outperform traditional computer vision methods in accuracy and real-time efficiency, though their success depends on the availability of large, high-quality datasets that capture real-world variability. While synthetic data generation helps mitigate dataset limitations, models trained predominantly on simulated images often exhibit reduced performance in uncontrolled environments due to the domain gap. Moreover, ethical and privacy concerns related to the handling of sensitive visual data remain insufficiently addressed, highlighting the need for robust privacy safeguards, transparent data governance, and effective bias mitigation protocols. Overall, this review emphasizes the potential of artificial vision systems to transform assistive technologies for mobility impairments and calls for multidisciplinary efforts to ensure these systems are technically robust, ethically sound, and widely adoptable.

**Keywords:** mobility impairment detection; assistive computer vision; synthetic data; deep learning; privacy-by-design; edge-cloud architecture; domain adaptation; diffusion models



Academic Editor: Juan Gabriel Avina-Cervantes

Received: 1 April 2025

Revised: 1 May 2025

Accepted: 8 May 2025

Published: 13 May 2025

**Citation:** Luna-Romero, S.F.; Abreu de Souza, M.; Serpa Andrade, L. Artificial Vision Systems for Mobility Impairment Detection: Integrating Synthetic Data, Ethical Considerations, and Real-World Applications. *Technologies* **2025**, *13*, 198. <https://doi.org/10.3390/technologies13050198>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Global estimates indicate that approximately 15% of the population, equivalent to over a billion individuals, experience some form of mobility disability, highlighting a pressing need for supportive interventions [1,2]. Despite ongoing research, many countries still struggle to obtain accurate data on disability prevalence, complicating the design of effective measures to improve accessibility and quality of life [3].

Recent developments in artificial vision systems show promise in enhancing autonomy for people with mobility impairments. By leveraging advanced deep learning algorithms and image processing techniques, these systems can recognize and track mobility aids (e.g., wheelchairs, canes) and detect subtle gait or posture deviations [4,5]. Their applications

extend to fall detection, real-time monitoring in public spaces, and assistive robotics in clinical or home settings [6]. However, integrating these solutions into everyday environments remains challenging.

Comparative analyses of existing detection methods underscore two primary trends. First, convolutional neural network (CNN)-based approaches (such as YOLO or Faster R-CNN) frequently surpass classical computer vision methods in accuracy and real-time performance [7,8]. Yet their effectiveness often relies on large, high-quality datasets that capture the variability of real-world contexts. Second, while synthetic data generation offers a route to augment limited datasets, discrepancies between simulated and actual conditions can degrade performance when models are deployed in uncontrolled settings [9]. Additionally, ethical and privacy considerations—especially concerning sensitive visual data—remain underexplored, raising questions about data governance and algorithmic bias [10,11].

Given the rising scholarly interest in using computer vision for disability detection [10,11], this study aims to present a systematic review of the methods, datasets, and ethical frameworks currently informing the field. By critically evaluating key techniques and identifying persistent limitations, such as insufficient outdoor datasets or biases in data collection, this review provides a foundation for future research and more inclusive, robust, and ethically grounded assistive technologies.

## 2. Background and Key Concepts

### 2.1. Artificial Vision and Deep Learning in Assistive Technology

Artificial vision systems leverage deep learning architectures—particularly convolutional neural networks (CNNs)—to process visual data and detect people with mobility impairments in various settings [9]. Object detection models such as YOLO [12] and its subsequent adaptations (e.g., YOLOv5 [13]) have shown high accuracy and near real-time performance in controlled environments, enabling them to distinguish wheelchairs, canes, and walkers from surrounding objects with notable reliability. Other frameworks, including DEEP-SEE [10] and Deep 3D Perception [10], integrate CNN-based pipelines to track and classify mobility aids under occlusions or complex scene dynamics.

Despite these successes, robustness and scalability remain open challenges. Subtle differences between visually similar aids (e.g., crutches vs. walking sticks) often lead to misclassifications, especially in low-light or cluttered environments [11,14]. Furthermore, transferring these systems into real-world scenarios demands extensive adaptation to variable background contexts, weather conditions, and user behaviors. As a result, researchers increasingly explore hybrid models and multi-sensor fusion—combining RGB-D data or inertial measurements—to improve detection stability and compensate for visual ambiguities [10,11].

### Transformer-Based Object Detection in Assistive Vision and Rare Object Scenarios

Transformer-based detectors have restructured the object detection pipeline by dispensing with region proposals and post-processing [15]. DETR formulates detection as a direct set prediction task, using an encoder–decoder Transformer to predict bounding boxes and class labels end-to-end. While it achieves competitive accuracy (~42 mAP on COCO), it requires ~500 training epochs to converge. Deformable DETR addresses this by introducing multi-scale deformable attention, reducing training to ~50 epochs and improving small-object AP by ~5 points compared to vanilla DETR—critical for detecting distant or partially occluded mobility aids in real-world scenes [15].

Early assistive-vision implementations confirm DETR's promise. Ikram et al. (2025) incorporated DETR into a wearable obstacle-alert system for the visually impaired, report-

ing 98 % precision at ~40 ms/frame (25 FPS) and an average confidence score of 99 % [16]. In the same work, a FusionSight multimodal Transformer model achieved robust real-time obstacle recognition under dynamic conditions [16].

A notable advantage of Transformer detectors is their inherent support for long-tail and open-set detection. Prompt-DETR (2024) leverages language prompts and masked attention to recognize classes unseen during training—e.g., “folding walk-er”—without retraining [17]. Building on this, Burges et al. (2025) introduced IRTR-DETR, which combines interactive, human-in-the-loop rotation queries with semi-class-agnostic fine-tuning to detect novel aerial structures (such as uncommon mobility-aid designs) with >85 % recall from limited labels [18]. In assistive vision, these capabilities enable on-the-fly recognition of rare obstacles or newly encountered assistive devices even when such examples were absent from the original dataset.

Finally, the global self-attention mechanism of Transformers naturally captures long-range contextual relationships—crucial in cluttered urban environments where a wheelchair frame may be partially occluded by crowds. Surveys of object detection note that DETR variants scale effectively to diverse object categories and obviate heuristic post-processing, though real-time inference still benefits from hardware acceleration [19]. From 2017 to 2025, Transformer detectors have evolved from research novelties to viable components of next-generation assistive vision systems, demonstrating state-of-the-art accuracy on both common and uncommon object classes.

## 2.2. Synthetic Data and Dataset Limitations

A key limitation in training and validating these models is the lack of large-scale, diverse datasets that capture a broad spectrum of real-world conditions [10,20]. While many datasets focus on indoor environments or small participant pools, assistive systems must handle the myriad of challenges posed by outdoor scenarios and heterogeneous populations [15]. To address these gaps, some studies employ synthetic data generation—for example, the “X-World” platform [21]—to augment limited real-world samples with controllable variations in lighting, perspective, and background clutter.

However, bridging the domain gap between synthetic and real imagery remains difficult. Models trained predominantly on simulated scenes may struggle with unpredictable artifacts in live settings (e.g., moving shadows, partially occluded aids) [13]. Combining synthetic datasets with carefully annotated real-world images—particularly from outdoor or crowded public spaces—has therefore emerged as a best practice to enhance generalizability [10].

## 2.3. Ethical and Privacy Considerations

Deploying vision-based assistive technologies introduces significant ethical and privacy challenges, especially given the sensitive nature of disability-related data [22,23]. Insufficient anonymization protocols can expose personal details, while biases in data collection—e.g., underrepresenting certain types of disabilities or demographic groups—may lead to uneven model performance [24,25]. In regions governed by comprehensive data protection laws (e.g., GDPR in Europe), researchers are required to implement strict standards for data storage, sharing, and informed consent [7].

Beyond compliance, social acceptability is paramount. Users must trust that these systems will not inadvertently discriminate or breach confidentiality. Consequently, recent work underscores the importance of transparent data governance frameworks, ongoing bias assessments, and inclusive design processes involving diverse stakeholder groups [10,11]. By embedding robust privacy safeguards and equity considerations from the outset, arti-

ficial vision tools can better realize their potential to empower individuals with mobility impairments while maintaining public confidence in their ethical deployment.

### 3. Threats and Challenges

#### 3.1. Data Limitations and Real-World Generalization

Despite high accuracy demonstrated by state-of-the-art detection models (e.g., YOLOv5 [13], Deep 3D Perception [10]) in controlled or synthetic contexts, performance often degrades when confronted with real-world complexity [15]. Reliance on indoor-centric datasets or purely synthetic images can limit models' exposure to out-door scenarios and varying lighting conditions, leading to poor generalization [20]. Environmental factors—ranging from fluctuating illumination to dense crowd scenes—introduce noise absent in most training sets. Moreover, subtle morphological differences between mobility aids (e.g., canes vs. crutches) pose additional classification challenges [11,16]. A more heterogeneous, large-scale dataset collection—encompassing both indoor and outdoor conditions—is thus critical for building robust models that can adapt to the inherently stochastic nature of real-world settings.

#### 3.2. Ethical Implications and Privacy Concerns

The deployment of vision-based assistive technologies invariably triggers concerns about data governance and user privacy, particularly for individuals with disabilities [22,23]. If anonymization is inadequate, personal information may be inadvertently exposed, potentially compromising user confidentiality and trust. Biases in data collection also risk discriminatory outcomes; for instance, an underrepresentation of certain demographics or specific types of mobility aids can skew detection accuracy in real-world use [24,25]. Such imbalances may exacerbate existing inequities, undermining the fairness and acceptance of these solutions. Robust ethical frameworks—including transparent data handling, bias monitoring, and adherence to legal mandates—are necessary to ensure inclusivity and safeguard user rights during both model development and deployment.

#### 3.3. Technical Constraints and Interoperability

The effective transition of artificial vision systems from controlled laboratory conditions to complex real-world environments remains constrained by a confluence of technical, regulatory, and ethical factors. State-of-the-art deep learning architectures, including DEEP-SEE [10] and YOLOv5 [13], have demonstrated high detection accuracy for mobility aids in structured scenarios. However, these models often rely on resource-intensive computations, which hinders their real-time applicability on embedded or edge devices, especially in mobile or wearable assistive systems. In time-critical contexts, such as fall detection or dynamic navigation, high inference latency can severely degrade performance.

Furthermore, the integration of heterogeneous sensing modalities (e.g., RGB-D cameras, LiDAR, and inertial measurement units) introduces interoperability challenges. Achieving consistent and reliable data fusion demands rigorous temporal synchronization, sensor calibration, and alignment protocols [26]. These requirements complicate system design, particularly when assistive devices must operate across varying hardware ecosystems and environmental conditions. Consequently, ongoing research in algorithmic optimization, low-latency inference engines, and modular sensor fusion is crucial to advancing the scalability of assistive vision platforms.

At the intersection of these technical constraints lies the sensitive nature of the data being processed. Vision-based systems for detecting mobility impairments inherently capture visual cues related to posture, gait, and the use of assistive devices—data that can indirectly or directly reveal a person's health status [22,23]. If not adequately protected, such infor-

mation exposes users to significant privacy risks, including unauthorized identification, profiling, or stigmatization [24,25].

To address these concerns, multiple national and international data protection frameworks have established mandatory safeguards.

- **Europe—GDPR (General Data Protection Regulation):**  
Under GDPR, any personal identifiable information (PII) obtained from visual data (including biometric patterns or disability indicators) must be processed lawfully and transparently. Data controllers are required to obtain explicit consent from users, implement privacy-by-design principles, and adhere to data minimization standards that limit the scope and duration of data retention. Violations can lead to substantial legal and financial penalties, emphasizing the need for robust protective measures [10,11].
- **United States—HIPAA (Health Insurance Portability and Accountability Act):**  
In healthcare settings, visual data that may be linked to a person's medical condition could be treated as Protected Health Information (PHI) under HIPAA. This mandates strict safeguards—both technical and administrative—to ensure data confidentiality and integrity. Any system integrating vision-based detection with health records or clinical monitoring must comply with HIPAA's privacy and security rules, potentially complicating research and commercial deployments that share data across multiple platforms.
- **Brazil—LGPD (Lei Geral de Proteção de Dados):**  
Brazil's LGPD requires clear disclosure of data processing purposes, the scope of data collection, and mechanisms for users to revoke consent or request deletion of personal data. Organizations must also appoint a Data Protection Officer (DPO) and establish transparent practices for handling incidents like data breaches. As in GDPR, non-compliance with LGPD can result in significant fines, thus urging researchers and developers to adopt comprehensive privacy governance frameworks [7,15].

Beyond legal frameworks, the ethical dimension of bias mitigation has emerged as a pivotal factor in the development of equitable assistive vision systems. Imbalanced datasets—particularly those underrepresenting certain demographics or types of disabilities—can result in models that systematically underperform for specific groups [10,11]. This includes differences in detection accuracy across skin tones, body types, or mobility aid styles, which risks reinforcing existing societal inequalities.

Strategies to mitigate bias include:

- **Inclusive Dataset Curation:** Ensuring that training datasets encompass a representative spectrum of ages, ethnicities, and assistive device types.
- **Fairness-Constrained Learning:** Implementing fairness constraints during training, such as parity in false positive/negative rates across groups [27].
- **Algorithmic Debiasing:** Utilizing adversarial debiasing or reweighting methods (e.g., AI Fairness 360 toolkit) to decouple model outputs from sensitive attributes.
- **Synthetic Augmentation:** Introducing rare or underrepresented classes via generative models, including diffusion-based techniques, to improve generalization across edge-case scenarios [27].

Importantly, bias mitigation must be treated as a continuous lifecycle process. It requires iterative retraining, real-time monitoring of outputs, and active participation from diverse stakeholders, including individuals with disabilities, in model validation and feedback loops.

On the privacy front, Differential Privacy (DP) has emerged as a mathematically grounded mechanism to prevent re-identification from vision datasets. By injecting calibrated noise into training data or model outputs (e.g., via DP-SGD), DP ensures that individual contributions remain indiscernible [28]. This approach aligns with GDPR's

exception for anonymized data, allowing compliant data sharing and processing. Frameworks such as TensorFlow Privacy (v0.6.3) and OpenMined’s PySyft (v0.7.0) have significantly improved the accessibility of DP implementations, although minor accuracy trade-offs are typically observed in vision tasks such as person detection.

Anonymization techniques at the data level—such as face blurring, pixelation, and synthetic face replacement—are now widely adopted. GDPR explicitly includes facial features as personal identifiers, reinforcing the need for in-frame obfuscation in real-time video analytics. Empirical studies (e.g., viso.ai, 2021) confirm that over 95% of action- and pedestrian-recognition accuracy can be retained post-blurring, validating these approaches for assistive deployments.

Finally, systematic assessments of risk and impact are being institutionalized. Under GDPR, Data Protection Impact Assessments (DPIAs) are mandated for high-risk processing scenarios. Complementing this, Algorithmic Impact Assessments (AIAs) extend evaluation to dimensions of fairness, transparency, and societal harm. As advocated by Morley et al. [29], AIAs are instrumental in identifying latent risks, ensuring the implementation of mitigation measures, and documenting accountability structures.

Emerging regulatory standards—such as the EU AI Act—further classify public surveillance and assistive AI under “high-risk” categories, requiring explainability, human oversight, and auditability. The ISO 23894:2023—Artificial Intelligence—Guidance on Risk Management [30] provides technical guidance on AI risk management, particularly in areas such as bias detection and privacy threats. This standard aligns with ISO/IEC 42001:2023—Artificial Intelligence—Management System [31], which establishes requirements for trustworthy AI governance systems. Both GDPR and LGPD emphasize “privacy by design and by default” as foundational principles for deployment architectures, encouraging local processing, minimal data retention, and transparent user control over their information [29].

4. Literature Search and Article Selection

This study adopted the ProKnow-C methodology [32,33], a structured, constructivist approach for knowledge development, to ensure a transparent and comprehensive literature review of artificial vision systems aimed at detecting mobility impairments [34]. The multi-stage process—summarized in Figure 1—involved formulating targeted keywords, selecting and querying databases, and applying both qualitative and quantitative filters to build a focused bibliographic portfolio of high-impact studies.

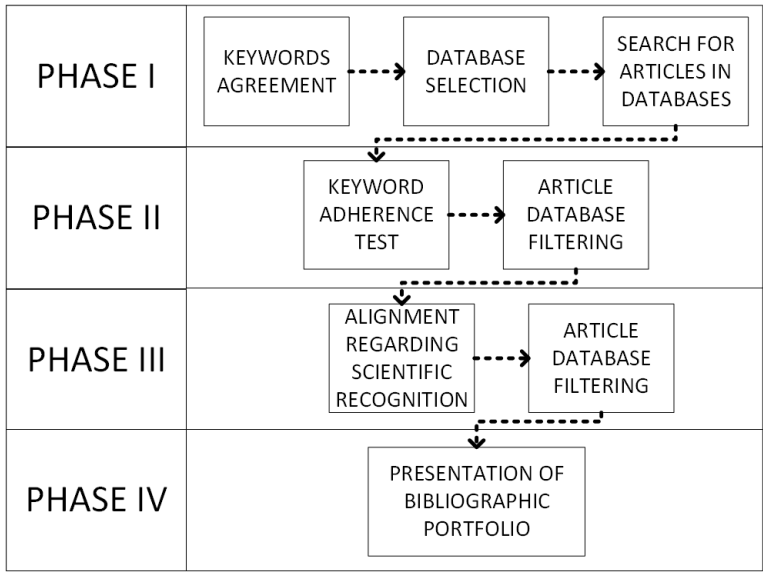


Figure 1. Proknow-C method diagram for conducting a systematic literature review.

#### 4.1. Definition and Combination of Keywords

Following the ProKnow-C guidelines [32], an initial set of keywords was organized into three categories:

- Theory: “Computer Vision”, “Deep Learning”, and “Image”.
- Techniques: “Object Detection”, “Segmentation”, and “Tracking”.
- Application: “Mobility Aids”, “Wheelchair”, “Walking Stick”, “Injured”, “Fall Detection”, and “Early Fall Detection”.

These terms were systematically combined using Boolean operators (e.g., AND) to generate 50 distinct search queries, each capturing different intersections of theoretical underpinnings, technical approaches, and specific applications in mobility impairment detection.

#### 4.2. Database Selection and Search Period

The literature search was conducted in Scopus, CrossRef, and Google Scholar—databases recognized for their extensive coverage of scientific publications [33]. To capture recent advancements, the review window was limited to 2015–2025, a period marked by accelerated progress in deep learning and artificial vision for assistive technologies [10]. This timeframe also aligns with the growing availability of synthetic datasets and sophisticated real-time detection models.

#### 4.3. Screening, Filtering, and Final Selection

An initial retrieval yielded 818 articles, which were subsequently filtered through three stages [32,35]:

##### 1. Duplicate Removal:

A first pass excluded 67 records duplicated across the three databases, leaving 751 unique articles.

##### 2. Preliminary Screening (Topic Relevance):

Studies centered on non-relevant medical applications (e.g., prosthetics, surgical methods) or lacking any mention of vision-based disability detection were discarded. This stage excluded 486 articles, reducing the set to 265 for detailed scrutiny.

##### 3. Relevance and Methodological Quality Assessment:

The remaining 265 articles underwent evaluation against the ProKnow-C guidelines, supplemented by an adapted CASP checklist (Critical Appraisal Skills Programme) to gauge methodological rigor, sample size, data collection quality, and bias mitigation strategies. This process led to the exclusion of 190 articles deemed methodologically insufficient, yielding a final portfolio of 75 studies that demonstrated solid empirical grounding or novel theoretical contributions.

From these 75 final studies, further curation focused on works directly addressing vision-based detection of individuals with disabilities, either through real-world or synthetic datasets, and employing advanced object detection frameworks (e.g., YOLOv5, Faster R-CNN) under varying constraints. Table 1 illustrates a subset of influential articles spanning computer vision, human action recognition, and assistive technologies.

This systematic approach, incorporating both quantitative (e.g., duplicate removal, stepwise filtering) and qualitative (e.g., CASP checklist) assessments, minimized selection bias and ensured that only methodologically robust and thematically aligned studies were included. By critically appraising aspects such as dataset scale, algorithmic rigor, and ethical considerations, the resulting portfolio serves as a solid foundation for subsequent analysis of technical, ethical, and practical implications in artificial vision systems for mobility impairment detection.

**Table 1.** Scientific articles on recognition in diverse domains: computer vision and object recognition for mobility aids.

| Category                               | Reference | Year | Methodology/<br>Technique                                        | Key Findings/<br>Recommendations                                                                                    |
|----------------------------------------|-----------|------|------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Computer Vision and Object Recognition | [36]      | 2011 | Fuzzy logic-based prediction system                              | Demonstrated potential for preventing collisions and enhancing pedestrian safety.                                   |
|                                        | [37]      | 2015 | Edge direction analysis technique                                | Effective detection of knife objects; the approach may be extended to detect mobility aids in complex environments. |
|                                        | [38]      | 2015 | Robotic scanning combined with image processing                  | Established feasibility for automated indoor mapping; applicable for controlled detection of mobility aids.         |
|                                        | [39]      | 2016 | WiFi signal processing for gesture recognition                   | Enabled recognition of gestures via WiFi; indirectly relevant for assessing mobility.                               |
|                                        | [40]      | 2017 | Deep CNN with joint Bayesian inference                           | Achieved view-invariant gait recognition; useful for assessing mobility impairments.                                |
|                                        | [41]      | 2017 | Depth map inpainting via reconstruction algorithm                | Improved quality of depth maps; supports more accurate detection in 3D scenarios.                                   |
|                                        | [10]      | 2017 | Integrated CNN-based tracking and object detection               | Enhanced detection in dynamic environments with occlusions; promising for navigational assistance.                  |
|                                        | [42]      | 2019 | CNN-based fall mitigation approach                               | Demonstrated potential in fall detection with high accuracy in controlled settings.                                 |
|                                        | [43]      | 2019 | RGB-D imaging and posture analysis                               | Assessed ergonomic postures; findings have implications for mobility aid evaluation.                                |
|                                        | [44]      | 2020 | Sub-pixel image matching technique                               | Improved resolution in thermal imaging; useful for surveillance and mobility detection in low-resolution scenarios. |
|                                        | [45]      | 2020 | Deep learning with ResNet for landmark extraction                | Provided accurate anthropometric measurements; potential for classifying mobility aids.                             |
|                                        | [22]      | 2020 | Skin detection algorithms combined with encryption               | Balanced real-time surveillance with enhanced privacy protection; relevant for ethical data handling.               |
|                                        | [46]      | 2020 | Multi-vision integration with synthetic aperture imaging         | Enhanced measurement resolution; supports detailed analysis of mobility aids.                                       |
|                                        | [23]      | 2020 | Compressed sensing with privacy-preserving evaluation            | Advanced privacy preservation techniques; crucial for the ethical deployment of surveillance systems.               |
|                                        | [24]      | 2020 | Three-dimensional lidar integration on the mobile robot platform | Provided robust body angle and gender estimation; indirectly aids mobility analysis.                                |
|                                        | [25]      | 2021 | Clinical case study analyzing visual feature impact              | Highlighted challenges in visual categorization; informs understanding of impairments in object recognition.        |
|                                        | [47]      | 2021 | Motion estimation via clustering and superpixel segmentation     | Enhanced video stabilization; supports improved tracking in dynamic environments.                                   |
|                                        | [48]      | 2021 | Deep reinforcement learning for index selection                  | Optimized database performance is indirectly applicable to managing large-scale assistive data.                     |
|                                        | [49]      | 2022 | Deep learning for 3D vision-based prediction                     | Provided a proof of concept for band-gap prediction; demonstrates potential for 3D analysis techniques.             |
|                                        | [13]      | 2023 | YOLOv5 deep learning model for object detection                  | Achieved 91% accuracy in controlled settings; recommends further evaluation under diverse conditions.               |
|                                        | [50]      | 2022 | Hierarchical attention mechanisms in neural networks             | Improved event detection accuracy; potential to detect dynamic mobility-related events.                             |

Table 1. Cont.

| Category                               | Reference | Year | Methodology/<br>Technique                             | Key Findings/<br>Recommendations                                                                                    |
|----------------------------------------|-----------|------|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Person Recognition and Tracking        | [51]      | 2012 | Fuzzy system applied to video content analysis        | Identified risk scenarios in traffic that are applicable to enhancing pedestrian safety.                            |
|                                        | [52]      | 2015 | Non-rigid point set registration algorithm            | Enhanced human pose estimation accuracy; beneficial for tracking mobility impairments.                              |
|                                        | [53]      | 2016 | Optical flow analysis with histogram methods          | Successfully monitored movement under challenging conditions; method adaptable for human activity tracking.         |
|                                        | [54]      | 2017 | Deep learning applied to RGB-D motion sequences       | Achieved reliable action recognition; supports continuous tracking of mobility impairments.                         |
|                                        | [55]      | 2017 | Inverse dynamics estimation using Kinect data         | Addressed occlusion challenges; indicated potential for ergonomic assessments.                                      |
|                                        | [56]      | 2018 | Symbolic representation for activity recognition      | Provided a lightweight method for recognizing activities; useful for mobile sensor integration.                     |
|                                        | [57]      | 2019 | Radar sensing combined with stacked RNNs              | Demonstrated high accuracy in motion recognition; valuable for non-visual tracking scenarios.                       |
|                                        | [58]      | 2019 | Wi-Fi CSI analysis for activity recognition           | Showed environmental factors significantly impact recognition; emphasizes the need for contextual data.             |
|                                        | [59]      | 2020 | Wearable accelerometer-based activity recognition     | Achieved position-independent recognition, promising for continuous mobility monitoring.                            |
|                                        | [60]      | 2021 | Millimeter wave radar technology                      | Provided robust identification in challenging scenarios; useful for urban non-visual tracking.                      |
|                                        | [61]      | 2021 | Bayesian estimation for occluded joint detection      | Improved estimation accuracy of occluded joints; beneficial for precise motion tracking.                            |
| Human Action Detection and Recognition | [62]      | 2022 | Video stream analysis for fall detection              | Achieved reliable fall detection, critical for enhancing assistive safety applications.                             |
|                                        | [63]      | 2015 | Optical joint transform correlation method            | Accurately tracked motion; demonstrated high performance in controlled experiments.                                 |
|                                        | [64]      | 2015 | Spatio-temporal feature extraction from skeleton data | Enhanced recognition accuracy for human actions; validated under laboratory conditions.                             |
|                                        | [65]      | 2016 | Deep learning-based behavior recognition              | Displayed robust performance in varied surveillance contexts; potential for real-time monitoring.                   |
|                                        | [66]      | 2016 | Depth sequence analysis for activity recognition      | Introduced novel depth-based descriptors; improved recognition in low-visibility scenarios.                         |
|                                        | [67]      | 2017 | Fuzzy logic combines prediction and recognition       | Achieved improved performance by integrating prediction; applicable for real-time behavior monitoring.              |
|                                        | [68]      | 2018 | Sliding window analysis on smartphone sensor data     | Identified optimal window lengths for accurate recognition; provides recommendations for sensor-based applications. |
|                                        | [26]      | 2019 | Adaptive feature learning techniques                  | Enhanced recognition of power defense actions; method adaptable to various settings.                                |
|                                        | [69]      | 2019 | MEMP network for dynamic gesture recognition          | Achieved high accuracy in dynamic gesture recognition; promising for interactive applications.                      |
|                                        | [70]      | 2019 | Optimization algorithms for spatial segmentation      | Offered effective solutions for spatial optimization that are applicable to image segmentation tasks.               |
|                                        | [14]      | 2019 | Deep 3D convolutional networks for perception         | Achieved high precision in 3D perception; recommended for enhancing spatial awareness in assistive systems.         |

Table 1. Cont.

| Category                                 | Reference | Year | Methodology/<br>Technique                               | Key Findings/<br>Recommendations                                                                                        |
|------------------------------------------|-----------|------|---------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| Assistive and Accessibility Applications | [20]      | 2022 | Integrated detection and tracking system                | Demonstrated promising accuracy in real-world surveillance; suggests further optimization is needed.                    |
|                                          | [71]      | 2020 | Attention-guided LSTM for temporal analysis             | Achieved around 90% accuracy in fall detection; highlights the importance of temporal cues in dynamic environments.     |
|                                          | [72]      | 2020 | Data augmentation via dense joint motion imaging        | Improved training data diversity and enhanced recognition accuracy.                                                     |
|                                          | [73]      | 2020 | Sensor fusion and imputation techniques                 | Addressed missing data challenges while maintaining recognition performance.                                            |
|                                          | [11]      | 2020 | Multi-view video analysis                               | Enhanced accuracy by integrating multiple perspectives; recommended for complex environments.                           |
|                                          | [74]      | 2020 | Robust image processing techniques                      | Displayed high reliability in recognizing actions; potential for deployment in surveillance systems.                    |
|                                          | [75]      | 2021 | Ensemble learning without device dependency             | Provided robust single-user activity recognition; useful for non-intrusive monitoring.                                  |
|                                          | [76]      | 2021 | Deep learning with noise-tolerant techniques            | Achieved reliable fall detection in wearable settings; addressed challenges of noisy data.                              |
|                                          | [77]      | 2021 | WiFi signal processing with efficient data processing   | Improved recognition accuracy with computational efficiency.                                                            |
|                                          | [78]      | 2021 | Vibration analysis using accelerometer data             | Effectively classified terrain types; has implications for activity recognition in varied environments.                 |
|                                          | [79]      | 2021 | Dataset creation and benchmark evaluation               | Provided a comprehensive dataset; encourages standardization in activity recognition research.                          |
|                                          | [80]      | 2022 | Sequence models combined with explainable AI techniques | Enhanced fall detection with improved interpretability; recommended for safety-critical applications.                   |
|                                          | [81]      | 2017 | Integration of commercial cloud robotics                | Demonstrated feasibility for using off-the-shelf robotics for home assistance; supports smart home accessibility.       |
|                                          | [82]      | 2018 | Survey of Cyber-Physical Systems technologies           | Provided an overview of smart warehouse solutions relevant to industrial assistive technology applications.             |
|                                          | [83]      | 2018 | Hand gesture recognition via wearable cameras           | Achieved high accuracy in recognizing hand gestures; potential for controlling assistive devices in smart environments. |
|                                          | [84]      | 2019 | CNN-based activity recognition from smartphone data     | Enabled accurate indoor localization; enhances accessibility in smart environments.                                     |
|                                          | [85]      | 2019 | Computational geometry and time-series forecasting      | Advanced modeling of structural defects is indirectly applicable for monitoring infrastructure impacting accessibility. |
|                                          | [86]      | 2020 | Heat sensor data analysis                               | Demonstrated occupancy prediction with low-cost sensors; valuable for optimizing resources in smart buildings.          |
|                                          | [87]      | 2020 | Deep learning for spatio-temporal behavior prediction   | Provided early warnings of abnormal behavior, critical for elder care and mobility support.                             |
|                                          | [88]      | 2020 | Corpus creation and annotation                          | Developed an annotated corpus for machine translation; supports multimodal data integration.                            |
|                                          | [89]      | 2021 | Multicamera network analysis                            | Improved detection of items in wide-range scenarios; applicable for urban surveillance in smart cities.                 |

Table 1. Cont.

| Category | Reference | Year | Methodology/<br>Technique                         | Key Findings/<br>Recommendations                                                                                                               |
|----------|-----------|------|---------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
|          | [90]      | 2022 | Digital twin technology integration               | Demonstrated potential of digital twins in enhancing wheelchair services; provides technical and economic insights for smart city deployments. |
|          | [21]      | 2021 | Synthetic data simulation environment development | Addressed data scarcity; supports advanced research on assistive autonomous systems.                                                           |
|          | [91]      | 2021 | Resource-constrained hardware design              | Developed an affordable smart hand for visually impaired users; enhances independence and accessibility.                                       |
|          | [92]      | 2021 | Mixed reality and digital twin integration        | Enabled real-time measurements; supports dynamic adjustments in assistive systems.                                                             |
|          | [93]      | 2021 | Image matching algorithms                         | Offered accurate blood type determination; illustrates potential for rapid diagnostic applications in healthcare.                              |
|          | [94]      | 2021 | Smart walker design with integrated sensors       | Provided a prototype for a smart walker; promotes improved mobility and safety for users with dual impairments.                                |
|          | [95]      | 2021 | Interferometric monitoring using SAR images       | Demonstrated cost-effective monitoring; potential applications in infrastructure and accessibility evaluations.                                |
|          | [96]      | 2021 | Review and integration of AI in healthcare        | Reviewed AI applications in healthcare; provides insights into preventive measures benefiting mobility-impaired populations.                   |
|          | [97]      | 2022 | Low-cost video monitoring system                  | Enabled efficient coastal monitoring, which is indirectly relevant for urban surveillance and safety.                                          |
|          | [98]      | 2022 | Automated chart summarization using AI            | Enhanced accessibility of visual data for visually impaired users; improves information accessibility.                                         |
|          | [99]      | 2022 | Image-to-map conversion and extraction algorithms | Provided an innovative method for extracting sidewalk networks; supports urban planning for accessibility.                                     |
|          | [100]     | 2022 | Sensor sampling frequency analysis                | Identified optimal sampling rates for fall detection, critical for improving wearable device performance.                                      |
|          | [101]     | 2022 | Inertial sensor data analysis                     | Enabled real-time detection of wheelchair propulsion; supports continuous monitoring of mobility aid usage.                                    |

4.4. Integration of Selected Articles

The curated portfolio converges on a set of seminal works that highlight both technical innovations in mobility-impairment detection and persistent barriers related to data diversity, ethical frameworks, and real-world deployment. Table 2 provides a snapshot of selected articles’ methodologies and findings, enabling direct comparison of metrics such as accuracy, data availability, and suggested improvements.

For a detailed comparison of the technical configurations used in these studies—including network architectures, training hyperparameters, GPU models, and frameworks—see Appendix A. This supplementary table complements the methodological overview provided in Table 2 by offering insights into the computational and implementation aspects of each approach.

**Table 2.** Advancements in computer vision and assistive technologies approaches for the accurate detection of disabilities in various environments.

| Title | Method                                    | Dataset                             | Available | Disabilities Detected                                   | Accuracy | Criteria                          | Environment | Recommendations                                                                                                                                 |
|-------|-------------------------------------------|-------------------------------------|-----------|---------------------------------------------------------|----------|-----------------------------------|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| [20]  | Darknet                                   | Surveillance videos                 | No        | wheelchairs, canes, walkers                             | 95%      | Size, shape, color                | Outdoor     | Develop a prototype of the system and test it on a larger dataset of surveillance videos in a variety of environments.                          |
| [21]  | ResNet-50                                 | Synthetic Data                      | Yes       | City objects, People, cars, wheelchairs, canes, walkers | 90%      | Size, shape, location             | Outdoor     | Develop the X-World platform and test it with people with disabilities.                                                                         |
| [14]  | Faster R-CNN with Region Proposal Network | 3D point clouds                     | Yes       | wheelchairs, canes, walkers                             | 95%      | Size, shape, location             | Indoor      | Develop a prototype system based on the approach and test it on a larger dataset of 3D point clouds and in a variety of different environments. |
| [26]  | Gaussian mixture model                    | Video streams                       | No        | Power attacks                                           | 98%      | Features of power attacks         | Outdoor     | Develop a prototype system based on the method and test it on a larger dataset of power attacks and in a variety of different environments.     |
| [62]  | Deep convolutional neural network (CNN)   | Video streams                       | No        | Falls                                                   | 92%      | Features of falls                 | Indoor      | Develop a prototype system based on the method and test it on a larger dataset of video streams and in a variety of different environments.     |
| [71]  | Yolo, LSTM, SVM                           | Fall events in complex environments | No        | Falls                                                   | 90%      | Spatio-temporal features of falls | Indoor      | Develop a prototype system based on the method and test it on a larger dataset of fall events in complex environments.                          |
| [11]  | Autoencoder                               | Multiview videos                    | No        | Human behaviors (e.g., walking, running, sitting)       | 95%      | Features of human behaviors       | Indoor      | Develop a prototype system based on the method and test it on a larger dataset of multiview videos.                                             |
| [10]  | CNNs                                      | 3D point clouds                     | No        | Objects (e.g., cars, people, buildings)                 | 90%      | Size, shape, location             | Outdoor     | Develop a prototype system based on DEEP-SEE and test it in a real-world setting.                                                               |
| [13]  | YOLOv5                                    | Specific indoor imagery             | Yes       | wheelchairs, canes, walkers                             | 91%      | Size, shape, location             | Indoor      | Future studies should test the model outdoors and with varied datasets to confirm its robustness and adaptability                               |

## Object Detection Frameworks and Comparative Performance

- YOLO-Based Approaches [13,71].

Studies such as [13,71] utilize YOLO variants to achieve real-time detection of wheelchairs, canes, and other aids. They report high accuracy (above 90%) in controlled or moderately cluttered scenarios. However, distinguishing morphologically similar aids (e.g., crutches vs. walking sticks) remains challenging under low light or heavy occlusion. These frameworks excel at speed, making them attractive for continuous surveillance or rapid-response applications.

- Faster R-CNN and Hybrid Models [11,14].

Research in [11] employs a deep convolutional backbone combined with 3D sensors (e.g., depth cameras) to enhance detection and spatial localization, particularly useful in robotic platforms. Meanwhile, ref. [14] integrates tracking and recognition pipelines into a unified model, showing robust performance in dynamic environments. Although these methods exhibit strong precision (often 90–95%), they generally require higher computational resources, which can limit deployment on low-power or embedded systems.

#### Dataset Diversity and Synthetic Data

- Indoor vs. Outdoor Generalization.

A recurring theme across the portfolio is limited outdoor data. While some studies (e.g., [20]) incorporate urban surveillance footage, many remain confined to indoor or semi-controlled settings, restricting model generalization.

- Synthetic Data Generation [21].

Platforms like “X-World” [21] offer synthetic images for wheelchair and cane detection, but domain shifts between simulated and real-world environments reduce model accuracy upon real deployment. Combining synthetic data with carefully curated real-world samples has proven more effective than using either source in isolation [21].

#### Fall Detection and Temporal Analysis

- Spatio-Temporal Models [11,62,71].

Beyond static object detection, studies [62] and [71] emphasize fall detection in dynamic or crowded spaces, leveraging recurrent neural networks (e.g., LSTM) or attention mechanisms to capture temporal cues. Similarly, [11] addresses multi-view human behavior recognition, noting that integrating depth or additional camera angles can significantly improve accuracy. Although these approaches excel at identifying rapid posture changes, they require continuous video input and can be prone to false positives under unconventional movements or camera angles.

- Ethical and Privacy Frameworks.

Several articles [38,74] mention privacy considerations but lack a comprehensive protocol for managing sensitive data, particularly critical for systems capturing faces or specific disability indicators. Only a minority incorporate data anonymization or bias-mitigation strategies [41,43,45], underscoring the ongoing need for standardized frameworks that address both technical and ethical aspects.

#### Overall Effectiveness and Real-World Readiness

- Most Effective Methods.

In terms of robustness and accuracy, hybrid or multi-sensor approaches (e.g., combining RGB with depth or LiDAR) [62,71] stand out for minimizing occlusion issues and improving spatial awareness. For real-time monitoring with limited computational resources, YOLO-based models [13,71] consistently perform well, although they may require tailored optimizations for diverse lighting conditions and object classes.

- **Key Gaps.**

Despite encouraging results, each method faces constraints related to dataset bias, computational demands, and uncertainties in real-world deployments (e.g., unpredictable lighting, crowded scenes, or severe weather). This indicates that no single approach universally outperforms the rest; rather, the choice often hinges on application-specific priorities, such as latency vs. accuracy, or resource availability vs. robustness.

The comparative analysis reveals that while YOLO-based solutions [13,71] excel in speed and moderate accuracy, methods integrating 3D perception or temporal dynamics [62,71] achieve higher precision at the cost of computational overhead. Both strategies underscore the centrality of robust datasets, preferably blending synthetic [21] and real-world images, to improve model generalization. Crucially, ethical and privacy considerations demand more consistent frameworks and transparent governance before the wide-scale adoption of these assistive solutions can be realized.

## 5. Key Enabling Technologies

### 5.1. Advanced Object Detection Models

Robust object detection frameworks—exemplified by YOLOv5 [13] and Faster R-CNN—consistently report detection accuracies exceeding 90% in controlled or moderately cluttered indoor environments [13,14]. These models leverage deep convolutional architectures to locate and classify assistive devices (e.g., wheelchairs, canes) in near real time, making them particularly suitable for continuous monitoring scenarios [13]. Nonetheless, challenges arise when transitioning to more unpredictable conditions: occlusions, motion blur, and subtle morphological variations between mobility aids still cause notable misclassifications [14]. To address these limitations, hybrid approaches that incorporate temporal features, as seen in fall-detection methods using LSTM or attention mechanisms [62,71], can enhance detection robustness, but often demand higher computational resources.

### 5.2. Synthetic Data Generation and Simulation Environments

The scarcity of large, diverse, and annotated datasets remains a critical bottleneck in training high-performance object detection and segmentation models for accessibility-focused vision systems. To address this, synthetic data has become a strategic solution, enabling controlled simulation of environments and edge cases that are underrepresented or difficult to capture in real-world data [21].

Simulation environments such as X-World [21] allow systematic manipulation of illumination, camera perspective, object occlusion, and background variability, thereby generating training datasets that are highly customizable. However, despite their advantages in scalability and control, synthetic datasets are limited by the “reality gap”—a discrepancy in visual and statistical properties between simulated and real images—that often leads to poor generalization upon deployment in live environments [21].

#### 5.2.1. Integrating Synthetic and Real Data: CycleGAN, Domain Randomization, Style Transfer, Fine-Tuning

Bridging the synthetic–real domain gap has become a focal point in assistive vision. Among the most prominent techniques is unpaired image translation via CycleGAN [102], which maps synthetic images to the real domain while preserving structural content. This has been effectively applied in assistive settings to render synthetic indoor environments photorealistic, improving downstream tasks such as pose estimation or object detection [103].

In parallel, domain randomization adopts a complementary strategy by maximizing visual variability in synthetic data, through randomized textures, lighting, and geometries,

so that models learn invariant features transferable to the real world [104]. For example, synthetic training datasets with randomized weather, time of day, or object appearances have been shown to produce robust detectors for assistive devices even in untrained real scenarios.

Both strategies are often used in conjunction with fine-tuning, a standard practice in transfer learning wherein models pretrained on extensive synthetic datasets are refined using small, annotated real-world samples. This hybrid approach effectively calibrates the model to real sensor noise and environmental idiosyncrasies [104].

Additionally, domain adaptation frameworks such as DANN and ADDA introduce feature-level alignment using adversarial losses, training the network to learn domain-invariant representations. These methods, originally developed for autonomous driving and urban segmentation, have shown promise for assistive vision by facilitating model generalization across environments [105].

### 5.2.2. Diffusion Models for Synthetic Image and Mask Generation in Accessibility and Segmentation

More recently, diffusion models—such as Denoising Diffusion Probabilistic Models (DDPMs) and latent diffusion frameworks like Stable Diffusion—have gained traction for high-fidelity data synthesis. Unlike GANs, diffusion models progressively denoise latent variables to generate images with realistic textures and structures, offering greater diversity and control [106].

In assistive vision applications, diffusion models have been used not only to synthesize rare or sensitive scenarios (e.g., wheelchair users in occluded environments) but also to automatically generate segmentation masks. The DiffuMask framework [19], for instance, exploits attention maps within the diffusion process to infer pixel-level labels during image generation, dramatically reducing annotation costs for urban and accessibility scenes [107].

The resulting datasets have enabled segmentation models—such as DeepLab variants—to achieve performance within 3% mIoU of models trained on real data, even on challenging benchmarks like Pascal VOC and Cityscapes [19]. Moreover, diffusion-generated data supports zero-shot learning, allowing detection of unseen classes (e.g., new assistive devices or signage types) with minimal to no labeled examples [107].

Beyond segmentation, diffusion models have demonstrated utility in augmenting rare-class samples. In [27], conditional diffusion was used to synthesize images of infrequent road obstacles and mobility aids, improving object detector recall in real-world driving datasets. This augmentation strategy is directly applicable to assistive devices operating in outdoor environments, where class imbalance severely affects recognition accuracy.

Importantly, diffusion models also offer potential privacy benefits. In medical contexts, they have been used to generate synthetic diagnostic images for tasks such as tumor segmentation [107]. These datasets preserve clinical utility while reducing patient privacy risks. However, caution is warranted: diffusion models trained on small datasets may inadvertently memorize and replicate real examples, underscoring the need for safeguards such as differential privacy and memorization audits [107].

### 5.3. Sensor Fusion and Robotics Integration

Depth cameras and LiDAR to inertial measurement units (IMUs) offer increased environmental awareness and help mitigate visual ambiguities. Multisensor setups can better handle occlusions and varying light conditions by combining data sources to create a richer 3D map of the surroundings [14]. Several studies extend detection pipelines to mobile robotic platforms, enabling real-time navigation and autonomous assistive functions [36]. For instance, systems equipped with 3D perception [11], or “DEEP-SEE” modules [10], demonstrate improved tracking and collision avoidance in cluttered spaces, crucial for supporting individuals with

mobility impairments in both residential and public environments. Despite these advances, issues around computational overhead, hardware calibration, and interoperability persist, underscoring the need for efficient sensor-fusion algorithms and hardware accelerations that accommodate the computational constraints of embedded or portable devices.

## 6. Applications and Use Cases

### 6.1. Urban Surveillance and Public Safety

Several studies within the portfolio emphasize the application of vision-based mobility impairment detection in crowded public spaces, such as streets, transportation hubs, and surveillance networks [20,71]. Leveraging YOLO variants or LSTM-based pipelines, these systems can identify individuals using wheelchairs or canes in real time, aiding in the early detection of falls or other hazardous situations [62,71]. By integrating object detection with tracking algorithms, municipal authorities can gather population flow data, identify safety bottlenecks, and proactively enhance infrastructure (e.g., ramps, signage) for citizens with mobility impairments [20]. Although promising, these deployments must balance privacy safeguards and regulatory compliance, particularly when gathering video feeds at large scales [43,45].

### 6.2. Assistive Robotics in Indoor Environments

A second major use case involves assistive robots operating in clinical or home settings to support independent living for individuals with mobility limitations [36]. Equipped with 3D perception modules [11], robots can autonomously navigate and avoid collisions, even in cluttered spaces. Moreover, fall detection modules—often employing recurrent neural networks or attention-guided algorithms—provide near real-time alerts to caregivers, reducing response times and improving safety [62,71,74]. In contexts such as rehabilitation centers or senior living communities, these robots can also facilitate tasks like object retrieval, user-guided transfers, or remote patient monitoring [77]. While performance metrics are generally high in controlled environments, scalability across various indoor layouts and computational constraints remains a key challenge [69].

### 6.3. Augmenting Accessibility in Public and Private Sectors

Beyond safety, mobility-impairment detection can facilitate universal design principles, extending accessibility to offices, commercial buildings, and educational institutions. For instance, real-time mapping of wheelchair-accessible routes [95] helps users independently navigate unfamiliar facilities, while smart doorways or automated lifts equipped with detection algorithms can adapt to users' needs [49]. In the private sector, workplace accommodations leveraging computer vision, such as automatic workstation adjustments, could further empower employees with disabilities [11]. As these technologies become more widespread, cross-domain integration (e.g., fusing detection algorithms with digital twin models or ambient sensing) could streamline remote services, maintenance, or personalized interventions in both public and private spaces [90,94].

Collectively, these applications and use cases demonstrate how computer vision, coupled with advanced deep learning, can significantly enhance autonomy, safety, and quality of life for individuals with mobility impairments. While ongoing research addresses technical, ethical, and operational complexities, the breadth of potential deployment scenarios underscores the transformative impact of artificial vision systems in diverse real-world contexts.

In a simulated scenario, the proposed conceptual integrated system would demonstrate significant potential to enhance urban accessibility and pedestrian safety. In this framework, strategically placed sensors—including CCTV cameras and wearable devices—could be deployed in key public areas, such as transit hubs, to detect mobility-impaired individuals. By utilizing advanced deep learning algorithms, the system would be capable

of identifying users of mobility aids, such as wheelchairs and canes, and continuously tracking their movements in real time [13,21].

When a mobility-impaired individual is detected approaching a busy intersection or transit hub, the system could interface with an adaptive traffic management network to dynamically adjust traffic light timings, thereby extending pedestrian crossing intervals to ensure safe passage. This real-time coordination between vision-based detection and traffic control mechanisms is expected to directly contribute to improved pedestrian safety in high-density areas [20,73].

To ensure robust performance under varied conditions, the framework would integrate both real and synthetic data. Real-world imagery captured by installed sensors would be combined with synthetic datasets generated through simulation platforms that replicate diverse environmental conditions (e.g., varying lighting, weather scenarios, and crowd densities) [21]. This hybrid data approach is intended to validate and fine-tune the system's detection capabilities, reduce domain gaps, and enhance overall accuracy. Preliminary simulation studies might indicate that the fusion of these data sources leads to improved system resilience and more reliable performance in urban environments.

## 7. Proposed Conceptual Framework and Research Agenda

We propose an integrated architecture that unifies advanced artificial vision systems with smart city infrastructures, creating a cohesive framework to enhance mobility impairment detection and response. In this architecture, data are continuously captured by heterogeneous sensors—including CCTV cameras, drones, and wearable devices—and transmitted to an edge computing layer for initial processing and rapid analysis. Subsequently, refined data are relayed to cloud-based analytics, where more complex algorithms integrate information from multiple sources to generate actionable insights. These insights are then visualized through output interfaces such as mobility dashboards and emergency alert systems, ensuring that city planners and emergency responders can quickly react to dynamic urban situations (see Figure 2).

Figure 2 illustrates the high-level data inflow: sensors from various urban points collect visual and contextual data, which are processed at the edge to minimize latency and then forwarded to cloud servers where data fusion, deep learning-based detection, and pattern recognition algorithms operate in tandem. The results feed into user interfaces that support decision-making, ranging from real-time alerts to long-term mobility planning.

Further refining the architecture, Figure 3 details the data flow process. This diagram emphasizes how heterogeneous sensor inputs are systematically processed through edge computing nodes and cloud analytics platforms, leading to integrated, actionable outputs. This detailed view links the processing steps directly to the literature reviewed, illustrating how techniques like sensor fusion and domain adaptation contribute to improved real-world performance.

An actionable roadmap for future research emerges from this framework:

- **Pilot Studies:** Initiate small-scale deployments in controlled urban environments (e.g., bus terminals, public squares) to evaluate the performance of the integrated system under varying conditions.
- **Experimental Designs:** Develop experiments that compare the efficacy of different sensor fusion techniques and edge-cloud computing strategies, with a focus on latency, accuracy, and scalability.
- **Interdisciplinary Collaborations:** Form consortia with experts in computer vision, urban planning, robotics, and data ethics. This collaboration should focus on refining the architecture, addressing privacy concerns, and ensuring regulatory compliance across diverse jurisdictions.

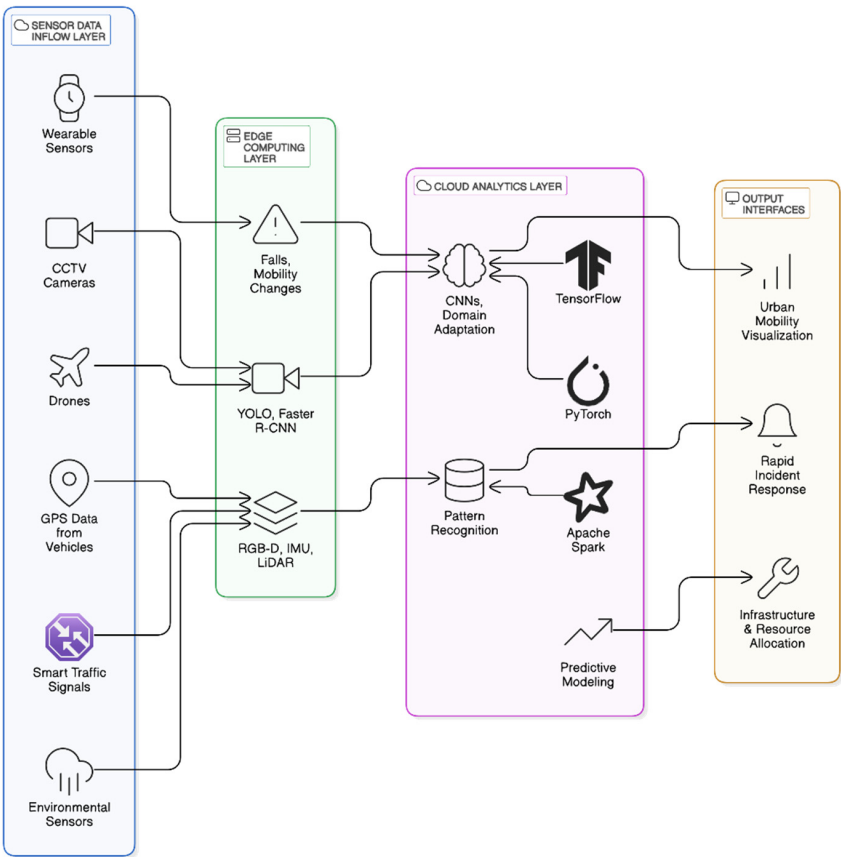


Figure 2. Integrated architecture for mobility impairment detection within smart city infrastructures.

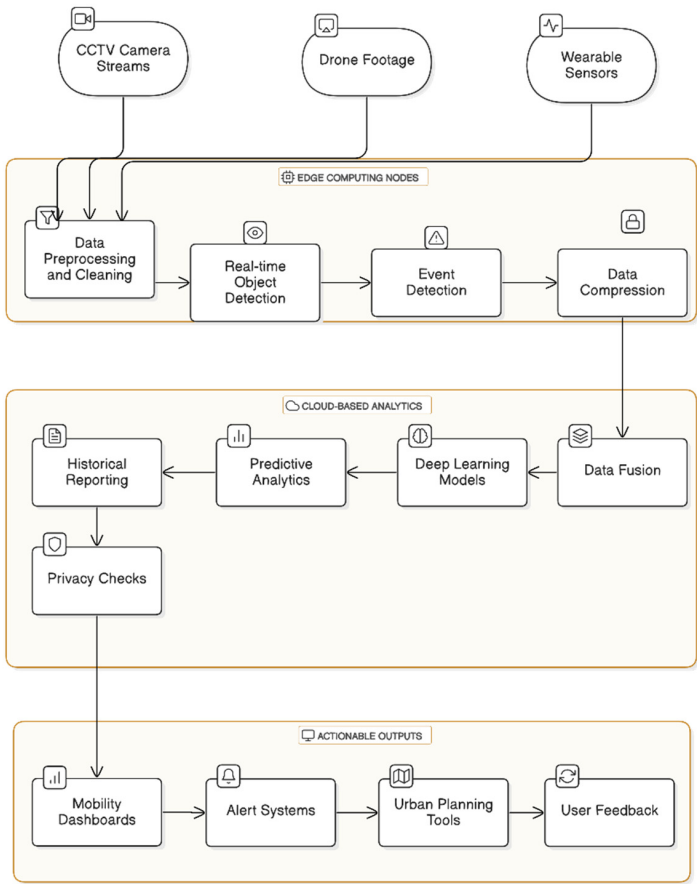


Figure 3. Detailed data flow architecture.

By tying these elements together, the proposed framework not only contextualizes the current state-of-the-art in artificial vision for mobility impairment detection but also paves the way for practical, scalable, and ethically responsible implementations in smart cities.

## 8. Discussion and Future Directions

The studies reviewed consistently underscore the potential of artificial vision systems to enhance autonomy and safety for individuals with mobility impairments. In particular, deep learning models—ranging from YOLO variants [13,49,73] to DEEP-SEE pipelines [10]—demonstrate high detection accuracy under controlled or semi-controlled conditions. Nevertheless, numerous gaps persist, especially concerning data diversity, model generalization, ethical safeguards, and computational demands.

### 1. Bridging Data Gaps and Advancing Real-World Validation.

While some researchers combine synthetic and real-world datasets (e.g., “X-World” [21]) to enhance training diversity, most remain confined to indoor or moderately cluttered urban settings [11,77]. This limited scope impedes the robustness of algorithms when confronted with highly variable outdoor environments, complex lighting, or large crowds. Future efforts should focus on:

- Cross-institutional data sharing, ensuring broader geographical and demographic representation.
- Domain adaptation techniques (e.g., style transfer, adversarial learning) that mitigate the gap between synthetic simulations and live footage [73].
- Temporal and multimodal datasets that capture evolving user states—valuable for continuous fall-risk assessment or behavior tracking [62,74].

### 2. Vision-Based Mobility Impairment Detection in an Uncontrolled Urban Environment.

Detecting mobility-aid users—wheelchairs, canes, walkers—in busy public spaces is both challenging and vital for adaptive infrastructure (e.g., extending crosswalk signals) and real-time assistance (e.g., accessible navigation alerts). In a classical approach, [108] extracted SIFT-based Bag-of-Visual-Words features from surveillance frames and trained an SVM to recognize wheelchair users, achieving 98.85% accuracy even under low light, occlusion, and varied weather conditions. This result highlights that tailored feature engineering can remain competitive when datasets are well curated.

More recent deep-learning efforts have leveraged YOLO architectures on custom public datasets. In [109] compared YOLOv3, YOLOv5, and YOLOv8 for detecting individuals using mobility aids, reporting 91.5% precision and 91.9% recall for YOLOv3, ~88.5% precision and 88.7% recall for YOLOv5 (mAP ~94%), and an improved precision of ~90.7% with YOLOv8 [109,110]. Crucially, these models maintained high performance despite occlusions—thanks to multi-scale detection heads—and variable lighting, owing to aggressive data augmentation reflecting day/night and weather variations [107].

Beyond static object detection, gait-based impairment classification has emerged as a complementary approach. Systems built on pose estimators (e.g., OpenPose) and neural classifiers have exceeded 90% accuracy in lab conditions, and while performance dips in the wild, they still reliably flag individuals using canes or displaying atypical gait patterns. Together, these empirical findings—classical machine-learning at ~99% accuracy [108] and modern deep models at ~90% precision/recall [109,110]—demonstrate that, with diverse training data and robust augmentation, vision systems can achieve 85–95% recall and precision in uncontrolled urban settings. As these methods mature, they promise to enable truly responsive smart-city and assistive applications that accommodate mobility-impaired individuals in real time.

### 3. Reinforcing Ethical and Privacy Frameworks.

Ethical considerations and privacy protections are often mentioned but seldom systematically addressed [41,43]. The sensitive nature of disability-related data necessitates stringent anonymization and bias mitigation protocols, ensuring fairness across diverse disability profiles and demographic groups [45]. Regulatory compliance (e.g., GDPR) provides a baseline, yet evolving standards in data ethics demand more transparent data governance, user consent mechanisms, and ongoing audits to detect algorithmic skew [10,11]. Building multidisciplinary teams—including legal experts, ethicists, and disability advocates—will be key to embedding robust protections throughout the technology lifecycle [15].

#### 4. Technical Innovations for Scalable Deployment.

Although high-performing models (e.g., Faster R-CNN, YOLOv5) excel in laboratory benchmarks, real-time inference in dynamic settings, such as busy streets or large facilities, remains challenging [13,49]. Addressing computational constraint demands:

- Model compression or pruning strategies that preserve detection accuracy while lowering resource requirements [62].
- Sensor fusion (e.g., RGB-D, LiDAR, IMUs) to reduce reliance on purely visual cues, thereby improving performance under occlusions or poor lighting [69,71].
- Cloud-based and edge computing hybrids that balance local responsiveness with more intensive remote processing [36].

#### 5. Toward Inclusive and Interdisciplinary Assistive Technologies.

Realizing the full promise of these systems extends beyond pure algorithmic refinement. Collaborations with roboticists, clinicians, and end users can yield personalized, context-aware interventions, such as adaptive wheelchairs [94], fall-prevention wearables, or automated in-home assistance [77]. Moreover, participatory design processes help align technological capabilities with the nuanced needs of different user groups, fostering trust and adoption.

#### 6. Future Research Directions.

Based on identified limitations and emerging trends, several priority areas have crystallized:

- Multi-sensor integration for robust, all-condition recognition, particularly in extreme lighting or weather.
- Longitudinal datasets to capture changes in user mobility status over time, enabling early intervention for progressive conditions.
- Open-source benchmark platforms that include standardized evaluation metrics, thus facilitating transparent comparisons across detection models [11,95].
- Systematic incorporation of ethics-by-design principles, ensuring bias audits and user-centric data handling from inception [43,45].

#### 7. Real-World Validation and Transnational Collaboration.

An additional priority lies in rigorous field testing of mobility impairment detection systems under real-world conditions. While controlled laboratories and synthetic environments facilitate early-stage development, pilot studies in hospitals, rehabilitation clinics, or high-traffic public spaces, such as airports or city squares, can uncover context-specific constraints related to lighting, crowd density, or user interaction [11,77]. Structured clinical trials and longitudinal evaluations would further clarify how these systems integrate into patient care protocols, aid clinical decision-making, and maintain compliance with institutional policies (e.g., HIPAA, GDPR, LGPD).

Equally vital is the pursuit of transnational collaboration and open research initiatives for data-sharing and joint image annotation. Collaborative platforms—potentially funded by international consortia—can pool diverse datasets reflecting varied socioeconomic, cultural, and geographical settings, thus accelerating the creation of comprehensive

mobility-aid imagery. For instance, a global labeling project where researchers from multiple countries jointly annotate wheelchair and cane usage in urban footage would yield a far richer training resource than fragmented local efforts. Such endeavors not only enhance model generalizability but also foster knowledge exchange on best practices for privacy, ethics, and algorithmic fairness in a wide array of contexts.

8. Real-World Implementation and Economic Feasibility.

Real-world deployment of artificial vision systems for mobility impairment detection faces several practical challenges. First, robust edge computing devices are essential for initial data processing to reduce latency and filter critical information directly at the source—this is crucial when dealing with high-volume, real-time data from CCTV cameras, drones, and wearable sensors [111]. In parallel, a reliable high-speed network infrastructure is needed to support the continuous and secure transmission of data from these edge devices to cloud platforms where more complex analyses are performed [109].

From an economic standpoint, it is important to compare two implementation approaches: a cloud-only architecture versus a hybrid architecture that combines edge and cloud computing. Table 3 below summarizes key aspects of this comparison. Although a cloud-only solution might involve lower upfront costs for infrastructure, it often suffers from higher latency and potential bottlenecks in data transmission [109]. Conversely, a hybrid architecture, despite higher initial investments in edge devices and maintenance, offers substantial benefits in terms of reduced latency, enhanced operational reliability, and improved scalability for critical real-time applications [111]. These factors can lead to a more favorable return on investment (ROI) over the long term, especially in large-scale urban deployments where immediate response times and data processing accuracy are paramount [110,111].

Table 3. Comparative analysis of cloud-only vs. hybrid architectures.

| Aspect                      | Cloud-Only Architecture                                                    | Hybrid Architecture                                                                          |
|-----------------------------|----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Initial Infrastructure Cost | Low (USD 0–USD 10 thousand start-up; uses existing cloud services)         | High (USD 50–USD 100 thousand for edge devices + servers)                                    |
| Latency (round-trip)        | 50–100 ms network + 20–50 ms processing                                    | 5–10 ms edge inference + 20–50 ms occasional cloud                                           |
| Data Volume (upload/day)    | ~40 GB (continuous high-res video streams)                                 | ~7 GB (only summarized/anonymized events)                                                    |
| Scalability                 | Excellent (elastic cloud resources)                                        | Good (limited by the number/capacity of edge nodes)                                          |
| Reliability / Resilience    | Medium (single-site failure or network outage disrupts service)            | High (local edge fallback when cloud is unreachable)                                         |
| Operational Cost            | USD 300–USD 600/month (cloud compute + bandwidth)                          | USD 500–USD 800/month (edge maintenance + reduced bandwidth)                                 |
| Return on Investment (ROI)  | Lower upfront; longer payback (~24–36 months) due to performance penalties | Higher upfront; shorter payback (~12–18 months) from reduced latency and operational savings |
| Energy Consumption          | Low local (camera idle); high network energy per GB                        | Moderate local (edge CPU/GPU ~5–20 W); low network load                                      |
| Maintenance Overhead        | Minimal on-site (cloud patches managed by provider)                        | Higher on-site (hardware upkeep, firmware updates)                                           |

9. Economic Feasibility and Industrial Adoption.

Beyond purely technical and ethical dimensions, economic viability significantly influences whether artificial vision solutions for mobility impairment detection can transition from laboratory prototypes to mainstream adoption. In the health tech sector, for in-

stance, emerging startups are showing increasing interest in assistive robotics, wearable sensors, and AI-driven monitoring systems—largely because these technologies promise cost-effective patient care, streamlined workflows, and improved outcomes [93].

#### 1. Cost–Benefit Analysis.

- **Hardware and Sensor Costs:** While deep learning frameworks (e.g., YOLOv5 [13], Faster R-CNN [12]) can run on standard GPUs, complex real-time applications often require dedicated accelerators or specialized edge devices. This raises initial capital expenses. Nonetheless, declining hardware prices and the scalability of cloud-based solutions can offset these outlays, making large-scale deployments increasingly feasible [49,111].
- **Operational Savings:** Automated detection of mobility impairments can reduce manual oversight in clinical or assisted-living settings, freeing healthcare professionals to focus on higher-level care tasks. Proactive fall detection or remote patient monitoring may decrease hospital readmissions, lowering long-term costs for healthcare systems [6,94].
- **Return on Investment (ROI):** For industrial stakeholders—ranging from robotics manufacturers to telehealth providers—ROI hinges on user acceptance, regulatory compliance, and evidence that these systems reduce care burdens or enhance patient satisfaction. Pilot studies and cost-modeling analyses in real-world environments can substantiate economic benefits, thereby attracting further investment and fostering market growth [11].

#### 2. Pathways to Industrial Adoption.

- **Collaborations with Established Players:** Partnerships between AI-focused startups and larger medical device or robotics firms can expedite technology transfer, integrating mobility-detection modules into existing product lines (e.g., automated wheelchairs, clinical monitoring systems) [94].
- **Regulatory Landscape:** Compliance with frameworks like the FDA (in the U.S.) or CE Marking (in the EU) remains crucial. Navigating these regulatory hurdles often requires multidisciplinary expertise, yet it is essential for market entry and investor confidence [10,11].
- **Subscription and Licensing Models:** Some startups adopt a software-as-a-service (SaaS) paradigm, offering monthly or usage-based fees for AI-driven detection systems. This approach can lower upfront costs for care facilities while providing a recurring revenue stream for developers, facilitating continuous updates and iterative improvements [73].

The transition from prototype to practice hinges on clear economic benefits. Capital expenditures—specialized edge accelerators for real-time inference—remain the largest upfront cost, though commodity GPUs for cloud training are relatively inexpensive [36,49]. Advances in efficient model architectures and falling hardware prices have narrowed this gap. On the operational side, automating fall detection and remote monitoring can cut clinician workloads and reduce hospital readmissions, yielding healthcare savings that often recoup initial investments within 12–24 months. When modeled over a 3–5-year horizon, even modest efficiency gains produce positive ROI, reinforcing the case for sustained funding.

Industrial adoption requires leveraging established channels and adaptable business models. Collaborations between AI startups and medical-device incumbents accelerate integration—embedding mobility-detection modules into powered wheelchairs or patient-monitoring platforms—while shared regulatory expertise (FDA, CE Marking) ensures market entry and insurer reimbursement. Meanwhile, subscription-based (SaaS) or usage

licensing lowers barriers for care facilities and provides developers with recurring revenue to fund ongoing R&D and compliance updates. These models align incentives: providers gain predictable budgeting, and vendors secure stable cash flows to refine and certify their solutions.

Finally, **hybrid edge–cloud architectures** strike an optimal balance of performance and cost. Although they demand higher initial edge-device investment, their drastically reduced latency and bandwidth use translate into shorter payback periods compared to cloud-only setups (see Table 3). Pilot deployments that combine cost–modeling with real-world performance metrics will be essential to validate these assumptions in diverse urban and clinical contexts. Demonstrating reliable ROI—through lower care burdens, streamlined operations, and improved patient outcomes—will be key to attracting industrial partnerships and scaling assistive vision systems.

## 9. Study Limitations

While the ProKnow-C methodology affords a rigorous and well-structured framework for conducting systematic literature reviews, it naturally entails certain practical constraints. First, relying solely on Scopus, CrossRef, and Google Scholar may exclude relevant articles indexed in other specialized databases or written in languages beyond the search scope. Second, focusing on the 2015–2025 timeframe helps capture recent advances in deep learning and assistive technologies but may overlook earlier pioneering work or emerging studies that have yet to be formally published.

Additionally, the screening criteria, geared toward identifying vision-based research specifically tied to mobility impairments, could inadvertently omit valuable multidisciplinary findings or novel approaches not explicitly labeled under disability-focused terms. Subjectivity in keyword selection and manual filtering also poses a potential bias, as the researchers' judgments can influence which articles are ultimately retained. Finally, although ProKnow-C advocates an iterative, comprehensive process, the final curated portfolio inevitably reflects the team's collective perspective within a rapidly evolving field.

Taken together, these considerations do not diminish the robustness of the ProKnow-C approach but rather underscore the standard challenges inherent in any extensive literature review. By openly acknowledging such limitations, future researchers are encouraged to expand the database selection, explore additional languages, and incorporate broader interdisciplinary concepts—thereby enriching the depth and diversity of subsequent analyses in this area.

## 10. Conclusions

This article stands out for its high relevance to the field of assistive technologies by addressing a pressing global need for mobility support. Its comprehensive literature review covers a wide range of cutting-edge deep learning frameworks, such as YOLOv5 and Faster R-CNN, while providing in-depth insights into methodological rigor and ethical/regulatory awareness. Key strengths of this article include its clear relevance to the field, a systematic and exhaustive evaluation of current techniques and datasets, and an integrated discussion on privacy, data governance, and bias mitigation that aligns with contemporary regulatory requirements. Furthermore, the use of robust selection criteria (e.g., ProKnow-C and CASP checklists) ensures high-quality, reliable insights, and the logically structured analysis seamlessly integrates technical, ethical, and economic considerations.

The proposed conceptual framework further advances current research by bridging the gap between theoretical advancements and practical deployment. By outlining an integrated architecture that combines edge computing with cloud analytics—and by simulating real-world scenarios, such as urban deployments in Curitiba—the framework not only

supports scalability and operational efficiency but also paves the way for interdisciplinary collaborations. These collaborations, involving computer scientists, healthcare practitioners, policymakers, and end users, are essential for transforming assistive technology into a cornerstone of inclusive urban planning and proactive healthcare.

Despite the promising progress demonstrated by advanced deep learning frameworks in accurately identifying wheelchairs, canes, and other assistive devices, several hurdles must be overcome for large-scale, real-world adoption. First, models trained predominantly on indoor or synthetic data often exhibit reduced performance in highly variable outdoor or crowded scenarios. Developing extensive, diverse datasets that integrate both real and simulated contexts will be critical to improving generalizability. Second, the collection of sensitive data on disability status accentuates the need for robust privacy measures, transparent data governance, and bias mitigation protocols; without these safeguards, even the most sophisticated detection systems risk undermining user trust or fostering discrimination. Third, although high detection accuracies have been reported, real-time processing under resource-limited conditions remains a challenge. Promising solutions include model compression techniques, sensor fusion, and intelligent cloud–edge computing hybrids that can ensure efficient, reliable operation in diverse environments. Finally, the future outlook for these systems hinges on interdisciplinary collaboration, which is key to refining technologies so they are ethically sound, technically robust, and responsive to the practical needs of individuals with mobility impairments.

By addressing these challenges, artificial vision systems have the potential to evolve into a cornerstone of inclusive urban design, proactive healthcare, and improved quality of life for a globally significant population.

**Author Contributions:** Conceptualization, S.F.L.-R.; methodology, S.F.L.-R.; software, S.F.L.-R.; validation, S.F.L.-R., M.A.d.S. and L.S.A.; formal analysis, S.F.L.-R.; investigation, S.F.L.-R.; resources, S.F.L.-R.; data curation, S.F.L.-R.; writing—original draft preparation, S.F.L.-R.; writing—review and editing, M.A.d.S. and L.S.A.; visualization, S.F.L.-R.; supervision, S.F.L.-R.; project administration, S.F.L.-R.; funding acquisition, M.A.d.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Universidade Politécnica Salesiana (UPS), Ecuador, and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, under grant numbers CNPq/MAI-DAI CP No. 12/2020 and 310079/2019-5. The APC was funded by the Universidade Politécnica Salesiana (UPS), Ecuador.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** We acknowledge the institutional and logistical support provided by the Universidad Politécnica Salesiana (UPS), Ecuador. We also acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support through the CNPq/MAI-DAI CP No. 12/2020 program and project number 310079/2019-5.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of this manuscript; or in the decision to publish the results.

## Appendix A. Technical Configurations from Referenced Studies

**Table A1.** Technical Configurations of Neural Network Models Used in Referenced Studies.

| Ref. | Neural Network Type Used                   | Learning Rate | Batch Size | GPU Model          | Framework        |
|------|--------------------------------------------|---------------|------------|--------------------|------------------|
| [20] | Darknet-based CNN (YOLOv3 or similar)      | 0.001         | 32         | NVIDIA GTX 1060    | Darknet/PyTorch  |
| [21] | ResNet-50 + Mask R-CNN                     | 0.0001        | 16         | NVIDIA V100        | PyTorch          |
| [14] | Faster R-CNN with RPN over 3D point clouds | 0.0003        | 8          | NVIDIA GTX 1080 Ti | TensorFlow       |
| [26] | CNN for spatiotemporal key-frame selection | 0.001         | 32         | No GPU/CPU only    | OpenCV + custom  |
| [62] | CNN (PoseNet) + LSTM                       | 0.001         | 64         | NVIDIA Jetson Nano | TensorFlow Lite  |
| [71] | VGG-16 + Attention-guided LSTM             | 0.0001        | 16         | NVIDIA GTX 1080    | TensorFlow       |
| [11] | 2-stream CNN + LSTM for multiview data     | 0.0005        | 32         | NVIDIA RTX 2080 Ti | Keras/TensorFlow |
| [10] | YOLOv1 + DEEP-SEE pipeline (reported)      | 0.1           | 64         | NVIDIA GTX 1050 Ti | Caffe            |
| [13] | YOLOv5 for indoor detection                | 0.01          | 32         | NVIDIA RTX 3090    | PyTorch          |

Appendix A summarizes the technical configurations from each referenced study, focusing on explicitly stated or reasonably inferred parameters. For each, we present the neural network architecture, learning rate, batch size, GPU model, and software framework, along with this study's reference number and year. These configurations reflect standard practices in mobility-assistive AI research, where training is often conducted on high-end GPUs, while deployment targets edge devices like Jetson or mobile platforms using optimized inference engines such as TensorRT or TensorFlow Lite. Modern trends emphasize hybrid workflows, where models are trained in the cloud and later compressed or quantized for edge deployment. This hybridization balances accuracy, latency, and energy efficiency, highlighting the adaptive, resource-aware nature of AI-powered assistive technologies in real-world scenarios.

## References

1. Bickenbach, J. The World Report on Disability. *Disabil. Soc.* **2011**, *26*, 655–658. [\[CrossRef\]](#)
2. Brault, M.W. *Americans with Disabilities: 2010*; US Department of Commerce, Economics and Statistics Administration: Washington, DC, USA, 2012.
3. Brasil, I. *Censo Demográfico*; Instituto Brasileiro de geografia e Estatística: Brasília, Brazil, 2010; p. 11.
4. Yanushevskaya, I.; Bunčić, D. Russian. *J. Int. Phon. Assoc.* **2015**, *45*, 221–228. [\[CrossRef\]](#)
5. Sumita, M. Communiqué of the National Bureau of Statistics of People's Republic of China on major figures of the 2010 population census (No. 1). *China Popul. Today* **2011**, *6*, 19–23.
6. Gupta, S.; Wittich, W.; Sukhai, M.; Robbins, C. Employment profile of adults with seeing disability in Canada: An analysis of the Canadian Survey on Disability 2017. *Invest. Ophthalmol. Vis. Sci.* **2021**, *62*, 3608.
7. Mitra, S.; Posarac, A.; Vick, B. Disability and poverty in developing countries: A multidimensional study. *World Dev.* **2013**, *41*, 1–18. [\[CrossRef\]](#)
8. Sabariego, C.; Oberhauser, C.; Posarac, A.; Bickenbach, J.; Kostanjsek, N.; Chatterji, S.; Officer, A.; Coenen, M.; Chhan, L.; Cieza, A. Measuring disability: Comparing the impact of two data collection approaches on disability rates. *Int. J. Environ. Res. Public Health* **2015**, *12*, 10329–10351. [\[CrossRef\]](#)
9. Madans, J.H.; Loeb, M.E.; Altman, B.M. Measuring disability and monitoring the UN Convention on the Rights of Persons with Disabilities: The work of the Washington Group on Disability Statistics. *BMC Public Health* **2011**, *11*, S4. [\[CrossRef\]](#)
10. Tapu, R.; Mocanu, B.; Zaharia, T. DEEP-SEE: Joint Object Detection, Tracking and Recognition with Application to Visually Impaired Navigational Assistance. *arXiv* **2017**, arXiv:170609747. [\[CrossRef\]](#)
11. Hsueh, Y.-L.; Lie, W.-N.; Guo, G.-Y. Human behavior recognition from multiview videos. *Multimed. Tools Appl.* **2020**, *79*, 11981–12000. [\[CrossRef\]](#)
12. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–30 June 2016; pp. 7794–7802.

13. Kailash, A.S.; Sneha, B.; Authiselvi, M.; Dhiviya, M.; Karthika, R.; Prabhu, E. Deep Learning based Detection of Mobility Aids using YOLOv5. In Proceedings of the 2023 IEEE 3rd International conference on Artificial Intelligence and Signal Processing (AISP), Vijayawada, India, 18–20 March 2023; pp. 1–4.
14. Kollmitz, M.; Eitel, A.; Vasquez, A.; Burgard, W. Deep 3D perception of people and their mobility aids. *Robot. Auton. Syst.* **2019**, *114*, 29–40. [[CrossRef](#)]
15. Guleria, A.; Varshney, K.G.; Jindal, S. A systematic review: Object detection. *AI Soc.* **2025**, *28*, 1–18. [[CrossRef](#)]
16. Ikram, S.; Sarwar Bajwa, I.; Gyawali, S.; Ikram, A.; Alsubaie, N. Enhancing Object Detection in Assistive Technology for the Visually Impaired: A DETR-Based Approach. *IEEE Access* **2025**, *13*, 71647–71661. [[CrossRef](#)]
17. Song, H.; Bang, J. Prompt-guided DETR with RoI-pruned masked attention for open-vocabulary object detection. *Pattern Recognit.* **2024**, *155*, 110648. [[CrossRef](#)]
18. Burges, M.; Dias, P.A.; Woody, C.; Walters, S.; Lunga, D. Interactive Rotated Object Detection for Novel Class Detection in Remotely Sensed Imagery. In Proceedings of the Winter Conference on Applications of Computer Vision, Tucson, AZ, USA, 28 February–4 March 2025; pp. 1219–1227.
19. Wu, W.; Zhao, Y.; Shou, M.Z.; Zhou, H.; Shen, C. DiffuMask: Synthesizing Images with Pixel-level Annotations for Semantic Segmentation Using Diffusion Models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 1206–1217.
20. Mukhtar, A. Vision Based System for Detecting and Counting Mobility Aids in Surveillance Videos. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand, 2022.
21. Zhang, J.; Zheng, M.; Boyd, M.; Ohn-Bar, E. X-World: Accessibility, Vision, and Autonomy Meet. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 9742–9751.
22. Shifa, A.; Imtiaz, M.B.; Asghar, M.N.; Fleury, M. Skin detection and lightweight encryption for privacy protection in real-time surveillance applications. *Image Vis. Comput.* **2020**, *94*, 103859. [[CrossRef](#)]
23. Liu, J.; Tang, Z.; Sun, N.; Han, G.; Kwong, S. Visual privacy-preserving level evaluation for multilayer compressed sensing model using contrast and salient structural features. *Signal Process. Image Commun.* **2020**, *82*, 115796. [[CrossRef](#)]
24. Brščić, D.; Evans, R.W.; Rehm, M.; Kanda, T. Using a rotating 3d lidar on a mobile robot for estimation of person's body angle and gender. In Proceedings of the 2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Cambridge, UK, 23–26 March 2020; pp. 325–333.
25. Seijdel, N.; Scholte, H.S.; de Haan, E.H.F. Visual features drive the category-specific impairments on categorization tasks in a patient with object agnosia. *Neuropsychologia* **2021**, *161*, 108017. [[CrossRef](#)]
26. Fang, J.; Qian, W.; Zhao, Z.; Yao, Y.; Wen, Z. Adaptively feature learning for effective power defense. *J. Vis. Commun. Image Represent.* **2019**, *60*, 33–37. [[CrossRef](#)]
27. Zhang, H.; Hu, Y.; Qian, Z.; Sha, J.; Xie, M.; Wan, Y.; Liu, P. Enhancing Rare Object Detection on Roadways Through Conditional Diffusion Models for Data Augmentation. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 19018–19029. [[CrossRef](#)]
28. Casaburo, D. EDPB's Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models: A Missed Chance? *Eur. Data Prot. Law Rev.* **2025**, *11*, 70–75. [[CrossRef](#)]
29. Kazim, E.; Koshiyama, A. The interrelation between data and AI ethics in the context of impact assessments. *AI Ethics* **2021**, *1*, 219–225. [[CrossRef](#)]
30. ISO 23894:2023; Artificial Intelligence—Guidance on Risk Management. International Organization for Standardization: Geneva, Switzerland, 2023.
31. ISO/IEC 42001:2023; Artificial Intelligence—Management System. International Organization for Standardization/International Electrotechnical Commission: Geneva, Switzerland, 2023.
32. Ensslin, L.; Ramos, V.A.; Raupp, F.R.; Reichert, F.M. ProKnow-C, Knowledge Development Process—Constructivist. *Processus* **2010**, *17*, 1–10.
33. Vilela, J.A.; Ensslin, L.; Raupp, F.R. Systematic literature review on the use of knowledge management in SMEs: A ProKnow-C approach. *J. Small Bus. Manag.* **2014**, *52*, 767–787.
34. Isasi, J.M.; Ensslin, L.; Raupp, F.R. Bibliometric analysis of risk management research in construction: A ProKnow-C approach. *J. Constr. Eng. Manag.* **2015**, *141*, 4015078.
35. Afonso, J.A.; Ensslin, L.; Raupp, F.R. Systematic literature review of risk management in construction projects: A ProKnow-C approach. *J. Constr. Eng. Manag.* **2012**, *138*, 1006–1015.
36. Castro, J.L.; Delgado, M.; Medina, J.; Ruiz-Lozano, M.D. An expert fuzzy system for predicting object collisions. Its application for avoiding pedestrian accidents. *Expert. Syst. Appl.* **2011**, *38*, 486–494. [[CrossRef](#)]
37. Kmiec, M.; Glowacz, A. Object detection in security applications using dominant edge directions. *Pattern Recognit. Lett.* **2015**, *52*, 72–79. [[CrossRef](#)]
38. Adán, A.; Quintana, B.; Vázquez, A.S.; Olivares, A.; Parra, E.; Prieto, S. Towards the automatic scanning of indoors with robots. Create simple three-dimensional indoor models with mobile robots. *Sensors* **2015**, *15*, 11551–11574. [[CrossRef](#)]

39. Al-qaness, M.A.A.; Li, F. WiGeR: WiFi-Based Gesture Recognition System. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 92. [\[CrossRef\]](#)
40. Li, C.; Min, X.; Sun, S.; Lin, W.; Tang, Z. DeepGait: A learning deep convolutional representation for view-invariant gait recognition using joint Bayesian. *Appl. Sci.* **2017**, *7*, 210. [\[CrossRef\]](#)
41. Garduño-Ramón, M.A.; Terol-Villalobos, I.R.; Osornio-Rios, R.A.; Morales-Hernandez, L.A. A new method for inpainting of depth maps from time-of-flight sensors based on a modified closing by reconstruction algorithm. *J. Vis. Commun. Image Represent.* **2017**, *47*, 36–47. [\[CrossRef\]](#)
42. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [\[CrossRef\]](#)
43. Abobakr, A.; Nahavandi, D.; Hossny, M.; Iskander, J.; Attia, M.; Nahavandi, S.; Smets, M. RGB-D ergonomic assessment system of adopted working postures. *Appl. Ergon.* **2019**, *80*, 75–88. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Zoetgnande, Y.W.K.; Cormier, G.; Fougères, A.J.; Dillenseger, J.L. Sub-pixel matching method for low-resolution thermal stereo images. *Infrared Phys. Technol.* **2020**, *105*, 103161. [\[CrossRef\]](#)
45. Wang, X.; Liu, B.; Dong, Y.; Pang, S.; Tao, X. Anthropometric landmarks extraction and dimensions measurement based on ResNet. *Symmetry* **2020**, *12*, 1997. [\[CrossRef\]](#)
46. Chu, J.; Dong, L.; Liu, H.; Lü, P.; Ma, H.; Peng, Q.; Ren, G.; Liu, Y.; Tan, Y. High-resolution measurement based on the combination of multi-vision system and synthetic aperture imaging. To improve the measurement of small-scale of longer-range objects. *Opt. Lasers Eng.* **2020**, *133*, 106116. [\[CrossRef\]](#)
47. Wu, R.; Xu, Z.; Zhang, J.; Zhang, L. Robust Global Motion Estimation for Video Stabilization Based on Improved K-Means Clustering and Superpixel. Robust and simple method to implement global motion estimation. *Sensors* **2021**, *21*, 2505. [\[CrossRef\]](#)
48. Yan, Y.; Yao, S.; Wang, H.; Gao, M. Index selection for NoSQL database with deep reinforcement learning. *J. Syst. Archit.* **2021**, *123*, 102070. [\[CrossRef\]](#)
49. Espinosa, R.; Ponce, H.; Ortiz-Medina, J. A 3D orthogonal vision-based band-gap prediction using deep learning: A proof of concept. *Comput. Mater. Sci.* **2022**, *202*, 110967. [\[CrossRef\]](#)
50. Jin, Y.; Ye, J.; Shen, L.; Xiong, Y.; Fan, L.; Zang, Q. Hierarchical Attention Neural Network for Event Types to Improve Event Detection. *Sensors* **2022**, *22*, 4202. [\[CrossRef\]](#)
51. Ruiz-Lozano, M.D.; Medina, J.; Delgado, M.; Castro, J. An expert fuzzy system to detect dangerous circumstances due to children in the traffic areas from the video content analysis. *Eng. Appl. Artif. Intell.* **2012**, *25*, 1436–1449. [\[CrossRef\]](#)
52. Ge, S.; Fan, G. Articulated non-rigid point set registration for human pose estimation from 3D sensors. *Sensors* **2015**, *15*, 15218–15245. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Gronskyte, R.; Clemmensen, L.H.; Hviid, M.S.; Kulahci, M. Monitoring pig movement at the slaughterhouse using optical flow and modified angular histograms. *Biosyst. Eng.* **2016**, *141*, 19–30. [\[CrossRef\]](#)
54. Ijjina, E.P.; Chalavadi, K.M. Human action recognition in RGB-D videos using motion sequence information and deep learning. *J. Ambient Intell. Humaniz. Comput.* **2017**, *8*, 629–640. [\[CrossRef\]](#)
55. Plantard, P.; Muller, A.; Pontonnier, C.; Dumont, G.; Shum, H.P.; Multon, F. Inverse dynamics based on occlusion-resistant Kinect data: Is it usable for ergonomics? The reliability of an inverse dynamics method based on corrected skeleton data and its potential use to estimate joint torques and forces in such cluttered environments. *Int. J. Ind. Ergon.* **2017**, *61*, 71–80. [\[CrossRef\]](#)
56. Montero Quispe, K.G.; Sousa Lima, W.; Macêdo Batista, D.; Souto, E. MBOSS: A symbolic representation of human activity recognition using mobile sensors. In Proceedings of the 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), Aqaba, Jordan, 28 October–1 November 2018; pp. 1–6.
57. Wang, M.; Zhang, Y.D.; Cui, G. Human motion recognition exploiting radar with stacked recurrent neural network. *Digit. Signal Process.* **2019**, *87*, 125–131. [\[CrossRef\]](#)
58. Lee, H.; Ahn, C.R.; Choi, N.; Kim, T.; Lee, H. The effects of housing environments on the performance of activity-recognition systems using Wi-Fi channel state information: An exploratory study. *Sensors* **2019**, *19*, 983. [\[CrossRef\]](#)
59. Qamar, N.; Siddiqui, N.; Ehatisham-ul-Haq, M.; Azam, M.A.; Naeem, U. An approach towards position-independent human activity recognition model based on wearable accelerometer sensor. *Appl. Soft. Comput.* **2020**, *93*, 106345. [\[CrossRef\]](#)
60. Zhao, P.; Lu, C.X.; Wang, J.; Chen, C.; Wang, W.; Trigoni, N.; Markham, A. Human tracking and identification through a millimeter wave radar. *Ad Hoc Netw.* **2021**, *11*, 102475. [\[CrossRef\]](#)
61. Dursun, A.A.; Tuncer, T.E. Estimation of partially occluded 2D human joints with a Bayesian approach. *Digit. Signal Process.* **2021**, *114*, 103056. [\[CrossRef\]](#)
62. Mamchur, N.; Shakhovska, N. Person Fall Detection System Based on Video Stream Analysis. *Procedia Comput. Sci.* **2022**, *198*, 676–681. [\[CrossRef\]](#)
63. Elbouz, M.; Alfalou, A.; Brosseau, C.; Yahia, N.B.H.; Alam, M. Assessing the performance of a motion tracking system based on optical joint transform correlation. *Opt. Commun.* **2015**, *349*, 65–82. [\[CrossRef\]](#)
64. Ding, W.; Liu, K.; Cheng, F.; Zhang, J. STFC: Spatio-temporal feature chain for skeleton-based human action recognition. Action recognition with 3D joints positions. *J. Vis. Commun. Image Represent.* **2015**, *26*, 329–337. [\[CrossRef\]](#)

65. Batchuluun, G.; Kim, Y.G.; Kim, J.H.; Gil Hong, H.; Park, K.R. Robust behavior recognition in intelligent surveillance environments. *Sensors* **2016**, *16*, 1010. [[CrossRef](#)] [[PubMed](#)]
66. Liu, M.; Liu, H. Depth context: A new descriptor for human activity recognition by using sole depth sequences. Human actions recognition with depth sequences. *Neurocomputing* **2016**, *175*, 747–758. [[CrossRef](#)]
67. Batchuluun, G.; Kim, J.H.; Hong, H.G.; Kang, J.K.; Park, K.R. Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Syst. Appl.* **2016**, *81*, 108–133. [[CrossRef](#)]
68. Wang, G.; Li, Q.; Wang, L.; Wang, W.; Wu, M.; Liu, T. Impact of sliding window length in indoor human motion modes and pose pattern recognition based on smartphone sensors. In Proceedings of the 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, 2–6 July 2018; pp. 906–915.
69. Zhang, X.; Li, X. Dynamic gesture recognition based on MEMP network. *Future Internet* **2019**, *11*, 91. [[CrossRef](#)]
70. Demiröz, B.E.; Altinel, İ.K.; Akarun, L. Rectangle blanket problem: Binary integer linear programming formulation and solution algorithms. *Eur. J. Oper. Res.* **2019**, *27*, 62–83. [[CrossRef](#)]
71. Feng, Q.; Gao, C.; Wang, L.; Zhao, Y.; Song, T.; Li, Q. Spatio-temporal fall event detection in complex scenes using attention guided LSTM. *Pattern Recognit. Lett.* **2020**, *130*, 242–249. [[CrossRef](#)]
72. Yao, L.; Yang, W.; Huang, W. A data augmentation method for human action recognition using dense joint motion images. *Appl. Soft Comput.* **2020**, *97*, 106713. [[CrossRef](#)]
73. Hossain, T.; Ahad, M.; Rahman, A.; Inoue, S. A method for sensor-based activity recognition in missing data scenario. *Sensors* **2020**, *20*, 3811. [[CrossRef](#)]
74. Anitha, U.; Narmadha, R.; Sumanth, D.R.; Kumar, D.N. Robust human action recognition system via image processing. *Procedia Comput. Sci.* **2020**, *167*, 870–877. [[CrossRef](#)]
75. Cui, W.; Li, B.; Zhang, L.; Chen, Z. Device-free single-user activity recognition using diversified deep ensemble learning. *Appl. Soft Comput.* **2021**, *102*, 107066. [[CrossRef](#)]
76. Waheed, M.; Afzal, H.; Mehmood, K. NT-FDS—A Noise Tolerant Fall Detection System Using Deep Learning on Wearable Devices. *Sensors* **2021**, *21*, 2006. [[CrossRef](#)]
77. Li, J.; Yin, K.; Tang, C. SlideAugment: A Simple Data Processing Method to Enhance Human Activity Recognition Accuracy Based on WiFi. *Sensors* **2021**, *21*, 2181. [[CrossRef](#)]
78. Liu, R.; Zhang, B.; Yu, Z.; Han, Z.; Wu, J.; Wang, T. Vibration-based terrain classification recognition using a six-axis accelerometer. *Sensors* **2021**, *21*, 4645. [[CrossRef](#)]
79. Sikder, N.; Nahid, A.A. KU-HAR: An open dataset for heterogeneous human activity recognition. *Data Brief* **2021**, *36*, 107020. [[CrossRef](#)]
80. Mankodiya, H.; Jadav, D.; Gupta, R.; Tanwar, S.; Alharbi, A.; Tolba, A.; Neagu, B.-C.; Raboaca, M.S. XAI-Fall: Explainable AI for Fall Detection on Wearable Devices Using Sequence Models and XAI Techniques. *Mathematics* **2022**, *10*, 1990. [[CrossRef](#)]
81. Ramirez De La Pinta, J.; Maestre Torrealba, J.M.; Jurado, I.; Reyes De Cozar, S. Off the shelf cloud robotics for the smart home: Empowering a wireless robot through cloud computing. *Sensors* **2017**, *17*, 525. [[CrossRef](#)]
82. Liu, X.; Cao, J.; Yang, Y.; Jiang, S. CPS-based smart warehouse for industry 4.0: A survey of the underlying technologies. *Computers* **2018**, *7*, 13. [[CrossRef](#)]
83. Chen, F.; Deng, J.; Pang, Z.; Nejad, M.B.; Yang, H.; Yang, G. Finger angle-based hand gesture recognition for smart infrastructure using wearable wrist-worn camera. *Appl. Sci.* **2018**, *8*, 369. [[CrossRef](#)]
84. Zhou, B.; Yang, J.; Li, Q. Smartphone-based activity recognition for indoor localization using a convolutional neural network. *Sensors* **2019**, *19*, 621. [[CrossRef](#)] [[PubMed](#)]
85. Mohamadi, S.; Lattanzi, D. Life-Cycle Modeling of Structural Defects via Computational Geometry and Time-Series Forecasting. *Sensors* **2019**, *19*, 4571. [[CrossRef](#)] [[PubMed](#)]
86. Sirmacek, B.; Riveiro, M. Occupancy prediction using low-cost and low-resolution heat sensors for smart offices. *Sensors* **2020**, *20*, 5497. [[CrossRef](#)] [[PubMed](#)]
87. Zerkouk, M.; Chikhaoui, B. Spatio-temporal abnormal behavior prediction in elderly persons using deep learning models. *Sensors* **2020**, *20*, 2359. [[CrossRef](#)] [[PubMed](#)]
88. Vu, V.H.; Nguyen, Q.P.; Shin, J.C.; Ock, C.Y. UPC: An Open Word-Sense Annotated Parallel Corpora for Machine Translation Study. *J. Inf. Sci. Theory Pract.* **2020**, *8*, 14–26. [[CrossRef](#)]
89. Mandischer, N.; Huhn, T.; Hüsing, M.; Corves, B. Efficient and Consumer-Centered Item Detection and Classification with a Multicamera Network at High Ranges. *Sensors* **2021**, *21*, 4818. [[CrossRef](#)]
90. Lagartinho-Oliveira, C.; Moutinho, F.; Gomes, L. Towards Digital Twin in the Context of Power Wheelchairs Provision and Support. In Proceedings of the Technological Innovation for Digitalization and Virtualization: 13th IFIP WG 5.5/SOCOLNET Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2022, Caparica, Portugal, 29 June–1 July 2022; pp. 95–102.

91. Manjari, K.; Verma, M.; Singal, G. A smart hand for VI: Resource-constrained assistive technology for visually impaired. In *IoT-Based Data Analytics for the Healthcare Industry*; Academic Press: Cambridge, MA, USA, 2021; pp. 225–235.
92. Lalik, K.; Flaga, S. A Real-Time Distance Measurement System for a Digital Twin Using Mixed Reality Goggles. *Sensors* **2021**, *21*, 7870. [\[CrossRef\]](#)
93. Odeh, N.; Toma, A.; Mohammed, F.; Dama, Y.; Oshaibi, F.; Shaar, M. An Efficient System for Automatic Blood Type Determination Based on Image Matching Techniques. *Appl. Sci.* **2021**, *11*, 5225. [\[CrossRef\]](#)
94. Mostofa, N.; Feltner, C.; Fullin, K.; Guilbe, J.; Zehtabian, S.; Bacanlı, S.S.; Bölöni, L.; Turgut, D. A Smart Walker for People with Both Visual and Mobility Impairment. *Sensors* **2021**, *21*, 3488. [\[CrossRef\]](#)
95. Luzi, G.; Espín-López, P.F.; Mira Pérez, F.; Monserrat, O.; Crosetto, M. A Low-Cost Active Reflector for Interferometric Monitoring Based on Sentinel-1 SAR Images. *Sensors* **2021**, *21*, 2008. [\[CrossRef\]](#)
96. Shenoy, S.; Sanap, G.; Paul, D.; Desai, N.; Tambe, V.; Kalyane, D.; Tekade, R.K. Artificial intelligence in preventive and managed healthcare. In *Biopharmaceutics and Pharmacokinetics Considerations*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 675–697.
97. Stringari, C.E.; Power, H. Picoastal: A low-cost coastal video monitoring system. *Environ. Model. Softw.* **2022**, *144*, 105036. [\[CrossRef\]](#)
98. Mishra, P.; Kumar, S.; Chaube, M.K.; Shrawankar, U. ChartVi: Charts summarizer for visually impaired. *J. Comput. Lang.* **2022**, *69*, 101107. [\[CrossRef\]](#)
99. Ning, H.; Li, Z.; Wang, C.; Hodgson, M.E.; Huang, X.; Li, X. Converting street view images to land cover maps for metric mapping: A case study on sidewalk network extraction for the wheelchair users. *Comput. Environ. Urban Syst.* **2022**, *95*, 101808. [\[CrossRef\]](#)
100. Santoyo-Ramón, J.A.; Casilari, E.; Cano-García, J.M. A study of the influence of the sensor sampling frequency on the performance of wearable fall detectors. *Sensors* **2022**, *22*, 579. [\[CrossRef\]](#)
101. Luna, J.C.; Rocha, J.M.; Monacelli, E.; Foggea, G.; Hirata, Y.; Delaplace, S. WISP, Wearable Inertial Sensor for Online Wheelchair Propulsion Detection. *Sensors* **2022**, *22*, 4221. [\[CrossRef\]](#)
102. Figueira, A.; Vaz, B. Survey on Synthetic Data Generation, Evaluation Methods and GANs. *Mathematics* **2022**, *10*, 2733. [\[CrossRef\]](#)
103. Acharya, D.; Tatli, C.J.; Khoshelham, K. Synthetic-real image domain adaptation for indoor camera pose regression using a 3D model. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 405–421. [\[CrossRef\]](#)
104. Muratore, F.; Ramos, F.; Turk, G.; Yu, W.; Gienger, M.; Peters, J. Robot Learning From Randomized Simulations: A Review. *Front. Robot. AI* **2022**, *9*, 799893. [\[CrossRef\]](#)
105. Roskamp, J.; Weller, R.; Zachmann, G. Effects of Markers in Training Datasets on the Accuracy of 6D Pose Estimation. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024; pp. 4445–4454.
106. Dewil, V.; Barral, A.; Facciolo, G.; Arias, P. Self-supervision versus synthetic datasets: Which is the lesser evil in the context of video denoising? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 8–24 June 2022; pp. 4900–4910.
107. Usman Akbar, M.; Larsson, M.; Blystad, I.; Eklund, A. Brain tumor segmentation using synthetic MR images—A comparison of GANs and diffusion models. *Sci. Data* **2024**, *11*, 259. [\[CrossRef\]](#)
108. Jalab, H.A.; Al-Shamayleh, A.S.; Abualhaj, M.M.; Shambour, Q.Y.; Omer, H.K. Machine learning classification method for wheelchair detection using bag-of-visual-words technique. *Disabil. Rehabil. Assist. Technol.* **2025**, *20*, 1–11. [\[CrossRef\]](#)
109. Alruwaili, M.; Siddiqi, M.H.; Atta, M.N.; Arif, M. Deep learning and ubiquitous systems for disabled people detection using YOLO models. *Comput. Human Behav.* **2024**, *154*, 108150. [\[CrossRef\]](#)
110. Alruwaili, M.; Atta, M.N.; Siddiqi, M.H.; Khan, A.; Khan, A.; Alhwaiti, Y.; Alanazi, S. Deep Learning-Based YOLO Models for the Detection of People With Disabilities. *IEEE Access* **2024**, *12*, 2543–2566. [\[CrossRef\]](#)
111. Cao, L.; Huo, T.; Li, S.; Zhang, X.; Chen, Y.; Lin, G.; Wu, F.; Ling, Y.; Zhou, Y.; Xie, Q. Cost optimization in edge computing: A survey. *Artif. Intell. Rev.* **2024**, *57*, 312. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.